Cosmology Part III

Enrico Pajer

Contents

Overview

The topics covered in these lecture notes will be as follows:

- A smooth expanding universe, FLRW, distances, constituents and thermal history
- Inflationary background
- Cosmological perturbation theory (CPT)
- Inflationary perturbations
- Cosmic Microwave Background (CMB)
- Large Scale Structures (LSS)

We use the shorthand notations: D for Dodelson's Modern Cosmology book [23], W for Weinberg's Cosmology book [68]. For example D 3 is Chapter 3 of Dodelson's book, while W appB is appendix B of Weinberg's book.

Check For Understanding (CFU) This a question that requires a short back-of-the-envelope calculation or a some reasoning.

Notation, units and conventions We use units in which $\hbar = c = k_b = 1$. Therefore energy is temperature and inverse time or inverse length. On the other hand, we will try to keep the reduced Planck mass explicit, $M_{\rm Pl} = (8\pi G_N)^{-1/2}$. Beware that some authors use $M_{\rm Pl}$ to indicate the "full" Planck mass $G_N^{-1/2} \simeq 1.2 \times 10^{19}$GeV. The necessary conversion factors can be added using dimensional analysis and

$$c = 3\times 10^8 \frac{\rm m}{\rm sec}\,, \quad {\rm pc} = 3.2\,{\rm lightyears}\,, \quad {\rm year} = \pi \times 10^7\,{\rm sec}\,, \tag{0.1}$$

$$\hbar c = 0.2\,{\rm eV}\,\mu{\rm m}\,, \quad M_{\rm Pl} \simeq 2.4\times 10^{18}{\rm GeV}\,. \tag{0.2}$$

We use the mostly plus signature $(-,+,+,+)$. Latin indices indicate space, $i, j, \dots = \{1, 2, 3\}$, while greek indices run over spacetime, $\mu, \nu, \dots = \{0, 1, 2, 3\}$. 3D vectors are in boldface, e.g. $\mathbf{k}$ and $\mathbf{x}$. Unless otherwise specified, all tensors are expressed in terms of the FLRW coordinates

$$ds^2 = -dt^2 + a^2 dx^2\,. \tag{0.3}$$

Standard derivatives are represented with a comma and covariant derivatives with a semi-column $T^{\cdots}_{\dots,\mu} \equiv \partial_\mu T^{\cdots}_{\dots}$, $T^{\cdots}_{\dots;\mu} \equiv \nabla_\mu T^{\cdots}_{\dots}$. Symmetrization and anti-symmetrization of a pair of indices is indicated with $(\dots)$ and $[\dots]$ respectively and is defined to have weight 1

$$A_{(\mu\nu)} \equiv \frac{1}{2}\left(A_{\mu\nu} + A_{\nu\mu}\right)\,, \quad A_{[\mu\nu]} \equiv \frac{1}{2}\left(A_{\mu\nu} - A_{\nu\mu}\right)\,. \tag{0.4}$$

My convention for the Fourier transform are

$$F(\mathbf{x}) = \int_{\mathbf{k}} \tilde{F}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad F(\mathbf{k}) = \int_{\mathbf{x}} \tilde{F}(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}, \quad \text{with} \quad \int_{\mathbf{k}} \equiv \int \frac{d^3\mathbf{k}}{(2\pi)^3}\,, \quad \int_{\mathbf{x}} \equiv \int d^3\mathbf{x}\,. \tag{0.5}$$

There are surprisingly many conventions for the name of variables in perturbation theory. In particular, Newtonian gauge is written as

$$
\begin{aligned}
ds^2 &\equiv -\left(1+2\Psi_D\right)dt^2 + a^2\left(1+2\Phi_D\right)dx^i dx^j \delta_{ij} &\quad (0.6)\\
&\equiv -\left(1+2\Phi_W\right)dt^2 + a^2\left(1-2\Psi_W\right)dx^i dx^j \delta_{ij} &\quad (0.7)
\end{aligned}
$$

in Dodelson's (D) or Weinberg's (W) notations. The conversion is $\Psi_D = \Phi_W$ and $\Phi_D = -\Psi_W$. In these notes, we use Dodelson's notation everywhere except where we keep the label W explicit.

Part I
The homogeneous Universe

1 The Expanding Universe

In this section we introduce the theoretical framework used to describe a universe that is homogeneous and isotropic on large scales. We begin with the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric, which captures the observed large-scale structure of spacetime. We then discuss the Einstein equations applied to such a background, leading to the continuity, Friedmann and acceleration equations that govern the cosmic expansion for different equations of state. Finally, we provide a brief overview of the role of symmetries and maximally symmetric spaces, which justify the assumptions underlying the FLRW description of our universe.

1.1 The Friedmann-Lemaitre-Robertson-Walker metric

The Cosmic Microwave Background (CMB) radiation, which describes the universe 370,000 years after the Big Bang, appears isotropic to a part in 10^5, as shown in Fig. 2. The distribution of galaxies on scales much larger than about a few Mega parsec (Mpc) appears homogeneous, as show in Fig. 3. The fractional size of inhomogeneities ranges from 10^{-5} on Hubble scales to $\mathcal{O}(1)$ at around 5 Mpc. These observations indicate that there exist a reference frame in our universe for which constant time hypersurfaces are approximately homogeneous and isotropic. Using the theorem reported in Sec. 1.1, the metric of our universe therefore must be approximated on large scales by

$$ds^2 = -dt^2 + a(t)^2 \gamma_{ij}(x) dx^i dx^j \,. \tag{1.1}$$

where $\tilde{g}_{ij}(x)$ is the metric of the maximally symmetric spatial 3D hypersurface, which we derive in the following. We recall that a maximally symmetric space is colloquially one for which every point is indistinguishable from any other point. A few comments are in order

- One can always re-define time as above to ensure that $g_{00}(t) = -1$. This choice is called *cosmological time*, or often simply time. As we will see in the next lecture, cosmological time corresponds to the proper time of observers at rest in the coordinates x^i.
- x are *comoving coordinates*. The word "comoving" here should be understood as opposed to "physical" coordinates. A spacelike comoving distance Δx^i is related to a physical distance $\Delta x_{\rm ph}$ by a factor of the metric. For the metric in (1.1) this is just:

$$\Delta x_{\rm ph} = \sqrt{\Delta x^\mu g_{\mu\nu} \Delta x^\nu} = \sqrt{a^2 \Delta x^i \gamma_{ij} \Delta x^j} = a|\Delta x| \,. \tag{1.2}$$

- The simplest possibility would of course be a constant $a(t)$, which could then be re-absorbed into the definition of x. But this is not what we observe. Since the early 20th century we know that on average all nearby galaxies recede from us at a speed proportional to their distance. This was originally pointed out by the influential work of Edwin Hubble. His famous plot, reported in Fig. 1, suggests the existence of a mathematical relation called the *Hubble law*

$$v \equiv \dot{x}_{\rm ph} = H_0 x \,, \tag{1.3}$$

  for some constant H_0 known as the *Hubble constant.* Using (1.2) the Hubble law gives

$$\dot{x}_{\rm ph} = \partial_t \left[a(t)|x| \right] = \dot{a}|x| = \frac{\dot{a}}{a} a|x| = \frac{\dot{a}}{a} x_{\rm ph} \stackrel{!}{=} H_0 x_{\rm ph} \tag{1.4}$$

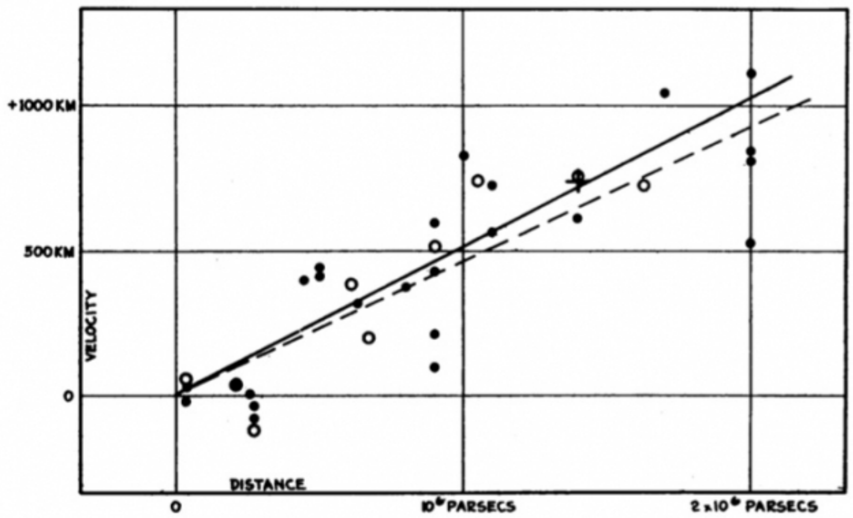


FIGURE 1
Velocity-Distance Relation among Extra-Galactic Nebulae.

Figure 1: Original plot of Hubble's data on the distance (horizontal axis) and velocity (vertical axis) of nearby galaxies.

where all time dependent quantities are evaluated today, $t = t_0$. We conclude that

$$\dot{a}(t_0)/a(t_0) = H_0 > 0\,, \tag{1.5}$$

We say that the universe is currently expanding.

We want to determine the explicit form of the 3D spatial metric $\tilde{g}_{ij}$. It is actually sufficient to find *one* maximally symmetric spatial metric, i.e. with all plus signature of a given curvature K, which must be the same at every point by symmetry. Because of the uniqueness theorem in Sec. 1.1, all other possible maximally symmetric spatial metrics with curvature K are related to this one by a change of coordinates. A very simple procedure is then to consider Euclidean space in one more dimension, i.e. $D = 4 + 0$, and derive the induced metric on the well-known constant curvature surfaces: the sphere ($K > 0$), the plane ($K = 0$) and the hyperboloid ($K < 0$). To minimise the use of indices and maximise transparency, I'll do this for a 2-dimensional surface embedded in 3 spatial dimensions. A generalization to any number of dimensions is straightforward and is left as an exercise.
Let us start with a 2-sphere in flat space

$$R^2 = x^2 + y^2 + z^2\,, \quad ds^2 = dx^2 + dy^2 + dz^2\,. \tag{1.6}$$

The induced metric is simply derived from the embedding

$$dz = -\frac{xdx + ydy}{\sqrt{R^2 - x^2 - y^2}}\,. \tag{1.7}$$

Going to "polar" coordinates, one obtains

$$\left\{ \begin{array}{l} x = \tilde{r}\cos\phi\,, \\ y = \tilde{r}\sin\phi \end{array} \right. \quad \Rightarrow \quad \begin{array}{rl} dl^2 = & \frac{d\tilde{r}^2}{1-\tilde{r}^2/R^2} + \tilde{r}^2 d\phi^2 \\ = & R^2\left[\frac{dr^2}{1-Kr^2} + r^2 d\phi^2\right]\,. \end{array} \tag{1.8}$$

Generalizing to our universe [Problem P.2.2] one finds

$$\begin{aligned} ds^2 &= -dt^2 + a^2\left[\frac{dr^2}{1-Kr^2} + r^2 d\Omega_2^2\right] && (1.9)\\ &= -dt^2 + a^2\left[d\chi^2 + f(\chi)d\Omega_2^2\right]\,, && (1.10)\\ f(\chi) &= \begin{cases} \sinh(\chi)^2 & K=-1\,(\text{open hyperbolic space})\,,\\ \chi^2 & K=0\,(\text{flat space})\,,\\ \sin(\chi)^2 & K=+1\,(\text{closed space or sphere})\,. \end{cases}\,. && (1.11)\end{aligned}$$

where K is the *spatial curvature* and

$$d\Omega_2^2 = d\theta^2 + \sin^2\theta d\phi^2\,. \tag{1.12}$$

Notice that $\chi \in \{0,\infty\}$ for open and flat universes, while $\chi \in \{\pi, 0\}$ for closed universes[1]. For a flat universe with $K = 0$, there is an ambiguity due to the rescaling $\{r, a\} \to \{\lambda r, \lambda^{-1}a\}$, which leaves the metric invariant. This is often fixed by imposing the additional condition that the scale factor $a(t)$ is set to unity today, namely

$$a(t_0) = a_0 = 1\,.$$

Unless otherwise specified we will always adopt this convention in these notes. For $K \neq 0$ this rescaling is fixed by normalizing $K = \pm 1$. It is sometimes convenient to have the metric in quasi-Cartesian coordinates as well, as opposed to spherical ones, namely[2]

$$\begin{aligned} ds^2 &= -dt^2 + a(t)^2\frac{dx^i dx^j\delta_{ij}}{(1+K\mathbf{x}^2/4)^2} && (1.13)\\ &= -dt^2 + a(t)^2 d\tilde{x}^i d\tilde{x}^j\left[\delta_{ij} + K\frac{\tilde{x}_i\tilde{x}_j}{1-K\tilde{\mathbf{x}}^2}\right]\,. && (1.14)\end{aligned}$$

There is no evidence of spatial curvature in our universe and current upper bound comes from a combination of CMB and Baryon Acoustic Oscillation (BAO) data

$$\left|\frac{K}{H_0^2}\right| = 0.000 \pm 0.005 \qquad (95\%\ \text{CL})\,.$$

For this reason, in these notes we will mostly focus on the flat case, $K = 0$. For future reference let us report the flat FLRW metric [homework P.1.3]

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t)dx^i\delta_{ij}dx^j\,, && (1.15)\\ &= a^2(t)\left[-d\tau^2 + dx^i\delta_{ij}dx^j\right]\,, && (1.16)\end{aligned}$$

where in the second line we introduced *conformal time*,

$$a d\tau \equiv dt\,,$$

[1] CFU: *What is the sum of the internal angles in a triangle in different geometries? What is the volume of a spatial slice? Show it is finite only for the sphere.*

[2] CFU: *Derive the relation between* x^i *and* $\{r, \theta, \phi\}$.

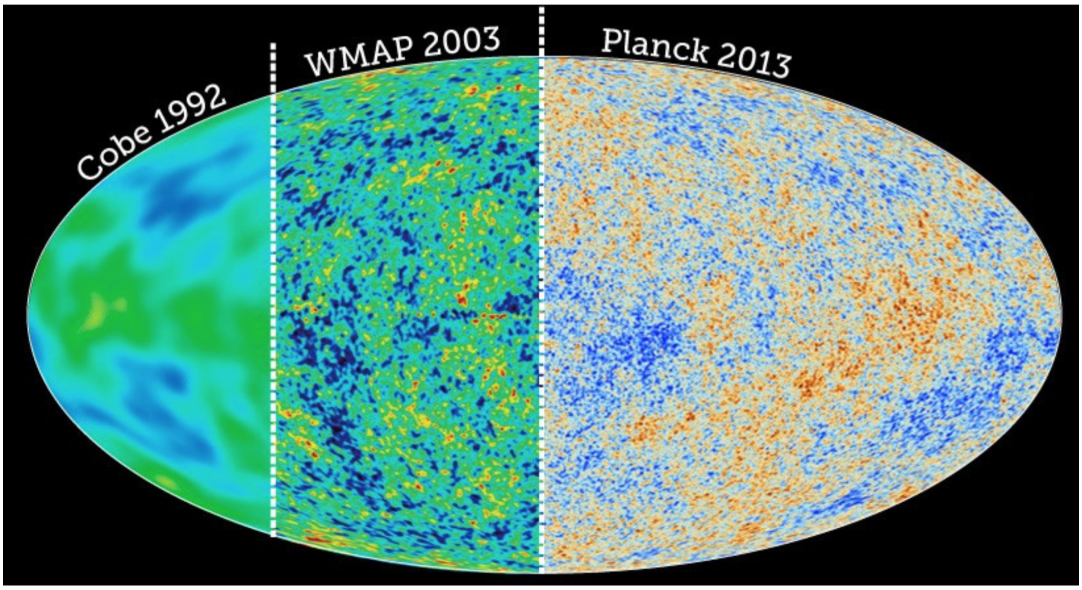


Figure 2: The temperature anisotropies in the CMB as seen by COBE with resolution $l_{max} \sim 20$ or $\theta_{min} \sim 9°$, by WMAP with resolution $l_{max} \sim 800$ or $\theta_{min} \sim 0.2° \simeq 12$ arcmin and by Planck with resolution $l_{max} \sim 2500$ or $\theta_{min} \sim 0.07 \simeq 4$ arcmin.

which makes manifest that the flat FLRW metric is proportional to the flat spacetime metric. In GR jargon we say that the FLRW metric is "conformally flat".

It is important to appreciate that FLRW space*time* is *not* maximally symmetric, since time translations and boost are broken by the generic time dependence of the scale factor $a(t)$. It is the constant-time 3d hypersurface that is maximally symmetric.
Recall that to define *covariant derivatives* that transform as tensor, we need to introduce the non-tensorial Christoffel symbols $\Gamma^{\mu}_{\sigma\nu}$

$$\Gamma^{\mu}_{\alpha\beta} \equiv \frac{1}{2} g^{\mu\gamma} \left(g_{\alpha\gamma,\beta} + g_{\beta\gamma,\alpha} - g_{\alpha\beta,\gamma} \right) , \tag{1.17}$$

These satisfy the useful properties

$$\Gamma^{\mu}_{\nu\rho} = \Gamma^{\mu}_{\rho\nu} , \quad \Gamma_{\mu\nu\rho} + \Gamma_{\nu\mu\rho} = g_{\mu\nu,\rho} , \quad \Gamma^{\mu}_{\mu\lambda} = \frac{1}{\sqrt{g}} \partial_\lambda \sqrt{g} . \tag{1.18}$$

The covariant derivatives are then given by

$$A^{\mu}_{;\nu} = \frac{\partial A^{\mu}}{\partial x^{\nu}} + \Gamma^{\mu}_{\sigma\nu} A^{\sigma} , \qquad A_{\mu;\nu} = \frac{\partial A_{\mu}}{\partial x^{\nu}} - \Gamma^{\sigma}_{\mu\nu} A_{\sigma} , \tag{1.19}$$

$$A^{\mu}_{\sigma;\nu} = \frac{\partial A^{\mu}_{\sigma}}{\partial x^{\nu}} - \Gamma^{m}_{\sigma\nu} A^{\mu}_{m} + \Gamma^{\mu}_{m\nu} A^{m}_{\sigma} , \qquad A_{\mu\sigma;\nu} = \frac{\partial A_{\mu\sigma}}{\partial x^{\nu}} - \Gamma^{\rho}_{\mu\nu} A_{\rho\sigma} - \Gamma^{\rho}_{\sigma\nu} A_{\mu\rho} . \tag{1.20}$$

For a flat FLRW metric, many Christoffel symbols vanish by the isotropy of FLRW spacetime[3].

[3]CFU: *What Christoffel symbols of FLRW vanish by symmetry? First, verify that* Γ *transforms as a tensor*

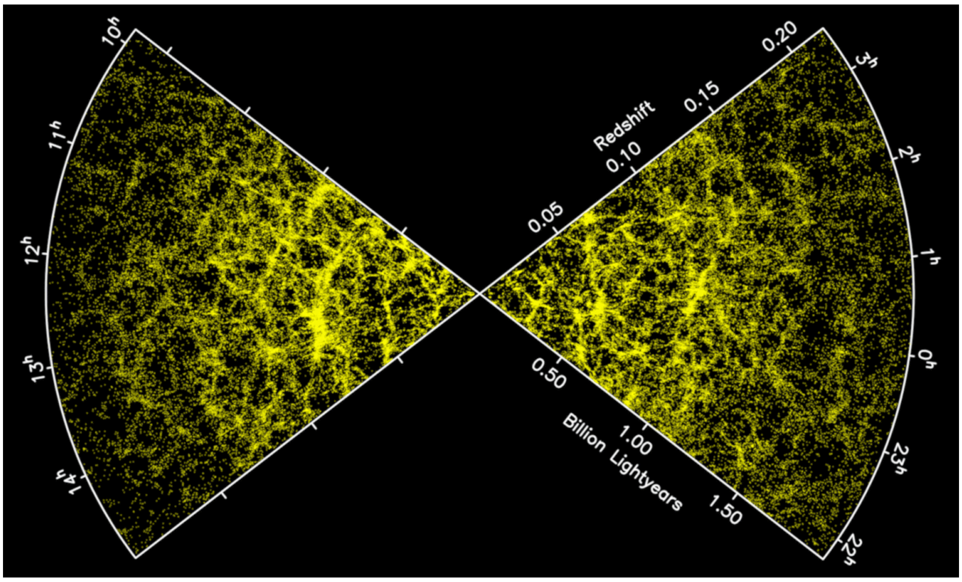


Figure 3: The distribution of galaxies as measured by the Sloan digital Sky Survey. The 3D dimension animated version is available here. In assessing homogeneity, keep in mind that more distant galaxies correspond to earlier time, when structures had had less time to grow.

Using the definition of γ_{ij} in (1.1), the non-vanishing Christoffel symbols are [homework **??**]

$$\Gamma^0_{ij} = Ha^2\gamma_{ij}\,, \quad \Gamma^j_{i0} = \Gamma^i_{j0} = H\delta_{ij}\,, \quad \Gamma^i_{jk} = \frac{1}{2}\gamma^{il}(\partial_j\gamma_{kl} + \partial_k\gamma_{jl} - \partial_l\gamma_{jk}), \tag{1.21}$$

For a flat FLRW metric $\partial_i\gamma_{jk} = 0$ and therefore $\Gamma^i_{jk} = 0$. We are now ready to see how space changes with time.

1.2 Te continuity equation and perfect fluids

In General Relativity (GR) the dynamical evolution of spacetime is dictated by Einstein's equations

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -\frac{1}{M_{\rm Pl}^2}T_{\mu\nu}\,, \tag{1.22}$$

for affine changes of coordinates $x'^{\mu'} = M^{\mu'}_{\mu}x^{\mu} + C^{\mu'}$ with C and M constant. This includes global rotations (both boosts and rotations) and spacetime translations. Only spatial rotations and spatial translations are symmetries of FLRW. The only tensor invariant under rotations is δ_{ij}. So $\Gamma^i_{00} = \Gamma^0_{0i} = 0$. Translations imply that $\Gamma = \Gamma(t)$. $\Gamma^0_{00} = 0$ is an "accident" of choosing the proper time of comoving observers t, instead of say τ or other $t' = f(t)$.

where we defined the Planck mass $M_{\rm Pl} = (8\pi G_N)^{-1/2}$ in natural units. Here $R_{\mu\nu}$ is a contraction of the Riemann tensor known as Ricci tensor,

$$R_{\mu\nu} \equiv R^{\rho}{}_{\mu\rho\nu} \,, \tag{1.23}$$

and $T_{\mu\nu}$ is the energy-momentum tensor. If the matter theory is described by an action S, then

$$T^{\mu\nu} \equiv \frac{2}{\sqrt{-g}} \frac{\delta S}{\delta g_{\mu\nu}} \,. \tag{1.24}$$

We will adopt the common notation to define the Einstein tensor as the left-hand side of Einstein's equations. Then we recall that, as consequence of the Bianchi identities of the Riemann tensor, which are reviewed around (A.31), the Einstein tensor satisfies

$$\nabla^{\mu} G_{\mu\nu} \equiv \nabla^{\mu} \left(R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \right) = 0 \,. \tag{1.25}$$

These equations are sometimes known as contracted Bianchi identities. Combining this result with the Einstein's equations we find that the energy-momentum tensor is covariantly conserved

$$\nabla_{\mu} T^{\mu\nu} = \partial_{\mu} T^{\mu\nu} + \Gamma^{\mu}{}_{\kappa\mu} T^{\kappa\nu} + \Gamma^{\nu}{}_{\kappa\mu} T^{\mu\kappa} = 0 \,. \tag{1.26}$$

Continuity equation Let us focus on the matter sector. The most general homogeneous and isotropic energy-momentum tensor takes the form

$$T^{\mu}{}_{\nu} = \mathrm{Diag} \left\{ -\rho, p, p, p \right\} \,, \tag{1.27}$$

where we will interpret ρ as the energy density, with units of E/L^3, and p as the pressure, with units $M/(T^2 L) = E/L^3$. The minus sign in $T^0_{\ 0}$ is related to the fact that T^{00} is the conserved energy density. Lowering indices with the metric gives

$$T_{00} = \rho \,, \qquad T_{0i} = 0 \,, \qquad T_{ij} = p g_{ij} \,. \tag{1.28}$$

Note that in cosmology we often make the assumption that the matter sector is well described by a perfect fluid, which we will discuss shortly. Here no such assumption has been made yet. The above form follows exclusively from symmetry.
As we just saw, Einstein's equations imply the covariant conservation of energy and momentum current, i.e. $T^{\mu\nu}{}_{;\nu} = 0$. The spatial components of this equation are trivial because of isotropy. Conversely, the time component plays a crucial role in cosmology:

$$\dot{\rho} + 3H \left(\rho + p \right) = 0 \,. \tag{1.29}$$

This is known as *continuity equation.* Sometime it is also referred to as "fluid equation", but we will avoid this terminology because it is valid also for non-fluid, as long as homogeneity and isotropy hold. To close the system of background equation we need to specify what type of matter we are considering. There are in principle a functional infinity of choices however most cosmological constituents are reasonably well described by the *equation of state*

$$p = w\rho \,, \tag{1.30}$$

with constant w. Let's emphasise that in general this equation of state is at best an approximation and realistic systems may have a much more complicated dependence on ρ or a time-dependent w. We will explore this in Sec. 6. For the simple linear equations of state, it is easy to solve the continuity equation to obtain $\rho(a(t))$,

$$\dot{\rho} + 3\frac{\dot{a}}{a}(1+w)\rho = 0 \quad \Rightarrow \quad \rho(t) = \rho(t_0)\left[\frac{a(t)}{a(t_0)}\right]^{-3(1+w)} . \tag{1.31}$$

For example:

- For non-relativistic matter, a.k.a. *dust*, most of the energy comes from the mass of the particles, with their kinetic energy giving small contributions. Hence the total energy is simply $E = mN$ with N the number of particles. Equivalently the energy density is $\rho = nm$ with $n = N/V$ the number density. Conversely the pressure p is proportional to the momentum mv of the particles moving at an average velocity v, hence $p \sim mvn$. Since in natural units $v \ll c = 1$, we find

  $$p \sim v\rho \ll \rho \quad \Rightarrow \quad w \sim 0 . \tag{1.32}$$

  In an FLRW universe dust evolves as $\rho \propto a^{-3}$. For an expanding universe, $\dot{a} > 0$, this is just the dilution with the volume one would expect from Newtonian intuition. While the approximation $w = 0$ will be sufficient most of the time, it is interesting to have an example of a small but finite value for w. For this, assume the particles are in thermal equilibrium at temperature T. The average velocity of a particle is obtained from $(1/2)mv^2 \sim T$ so that non-relativistic particles with $v \ll 1$ require $T \ll m$. Then, by the ideal gas law we know that the pressure obeys $pV = Nk_BT$, which in natural units $k_B = 1$ reads $p = nT$. From this we obtain $w = p/\rho = T/m$, which is indeed small. For more details see discussion around (6.26).

- For relativistic matter, or *radiation*, we have that energy and momentum are of the same order. Explicit calculation shows that $p = \rho/3$ or

  $$w = 1/3 \tag{1.33}$$

  (see around (3.1) for a detailed derivation). Expansion leads to the dilution $\rho \propto a^{-4}$. The extra factor of $1/a$ can be understood intuitively as the redshifting of energy in an expanding universe.

- For a cosmological constant, a.k.a. *vacuum energy*, $T_{\mu\nu} = \Lambda g_{\mu\nu}$ and therefore $p = -\rho$ or $w = -1$. As the name suggests, in an FLRW universe $\rho \propto a^0 \sim$ *const*.

Fluids A relativistic *perfect fluid* is defined as "as a medium for which at every point there is a locally inertial Cartesian frame of reference, moving with the fluid, in which the fluid appears the same in all directions." (see B.10 of [68]). In the comoving local inertial frame, indicated by $\doteq$, the energy-momentum tensor must be diagonal and isotropic: $T^{\mu}{}_{\nu} \doteq (-\rho, p, p, p)$. This is the same form we encountered above imposing homogeneity and isotropy. By boosting with a velocity u^μ, which is a timelike vector $u_\mu u^\mu = -1$, one finds the covariant form of the energy-momentum tensor

$$T^{\mu\nu} = (\rho + p)\, u^\mu u^\nu + g^{\mu\nu} p . \tag{1.34}$$

Here the energy density ρ and the pressure p are covariant scalars, while u^μ is a covariant vector. Conversely, if you are given some $T_{\mu\nu}$ in a spacetime with metric $g_{\mu\nu}$, you can check whether it is a perfect fluid or not by finding a solution u^μ to the following equation

$$\left(\Lambda^{-1}\right)^{\mu}_{\ \rho} T\Lambda(u) = \mathrm{Diag}(-\rho, p, p, p)\,, \tag{1.35}$$

where $\Lambda = \Lambda(u)$ is a Lorentz transformation with velocity u^μ and $\mathbf{T} = T_{\mu\nu}$ If $T_{\mu\nu}$ is that of a perfect fluid, then you can derive ρ, p and u^μ from (see Prob. **??**)

$$\rho \equiv \frac{1}{4}\left(\sqrt{12T_{\mu\nu}T^{\mu\nu} - 3T^2} - T\right)\,, \qquad p = \frac{1}{3}(\rho + T)\,, \qquad u_\mu u_\nu \equiv \frac{T_{\mu\nu} - g_{\mu\nu}p}{\rho + p}\,, \tag{1.36}$$

where $T \equiv T^{\mu\nu}g_{\mu\nu} = T^\mu_\mu$.

The energy-momentum tensor is again covariantly conserved $T^{\mu\nu}_{;\nu} = 0$. This can be seen as the conserved Noether current for the diffeomorphism invariance of the matter action (see e.g. Sec. 19.6 of [9]). Currents of gauge transformations (diffeomorphism are gauged by gravity) are identically conserved, and, in fact, $T^{\mu\nu}_{;\nu} = 0$ follows directly from Einstein Equations (1.22). In general, an equation of state $p = p(\rho)$ is necessary to close the system of equations. The extension to imperfect fluids is nicely discussed in B.10 of [68] and [64].
For the discussion of the Cosmic Microwave Background (CMB) and Large Scale Structures, we will have to consider more general "imperfect" fluids, with additional contributions to $T_{\mu\nu}$, that are organised in a derivative expansion as

$$T^{\text{imp.}} \sim T^{\text{perfect}}_{\mu\nu} + \sum_{n,m} d^n \nabla^n u^m \tag{1.37}$$

with some length scale d. The theory of general fluids, namely *hydrodynamics*, should then be thought of as a large scale effective theory, defined as an expansion in d/L, where L is the typical size of spatial variations in the fluid and in the classical examples d is the mean free path of the microscopic constituents.

If the fluid carries some *conserved charge* N, such as for example the number of particles, in the rest frame of the fluid one expects a charge density $n \equiv N/V$ for some small volume V that is conserved $\dot{n} \doteq 0$. To describe the conservation of n in any other frame, we must then have the covariant expression

$$\left(u^\mu n\right)_{;\mu} = 0\,, \tag{1.38}$$

which indeed reduced to $\dot{n} \doteq 0$ for $u^\mu = \{1, 0, 0, 0\}$. Since most processes in the universe are approximately adiabatic the total entropy of the universe is approximately conserved. An important covariantly conserved quantity is therefore the *entropy density* $s = S/V$, with entropy current su^μ. In general one finds

$$s = \frac{\rho + p - \mu n}{T}\,, \tag{1.39}$$

with n the number density of particles.

Charge conservation How does a conserved charge density depend on time in an FLRW universe? Isotropy implies that the normalised velocity defined in (A.43) and associated to the associated conserved current must take the form $u^\mu = \{1, \vec{0}\}$. Then, the conservation of the current $(nu^\mu)_{;\mu} = 0$ reduces simply to

$$\dot{n} + 3Hn = 0\,, \tag{1.40}$$

which admits the solution (see P.1.1)

$$n(t) = \left[\frac{a(t_0)}{a(t)}\right]^3 n(t_0) \propto \frac{1}{a^3}\,. \tag{1.41}$$

1.3 Friedmann equations

Now that we have some idea of what matter does as function of a, we can turn our attention to how the universe expands. To this end, let's solve the Einstein Equations (EEs) for an FLRW metric. Using the definition of the Riemann and Ricci tensors in (A.28), and the FLRW metric (1.13), a lengthy but straightforward computation in Cartesian coordinates shows that the only non-vanishing components are

$$R_{00} = 3\frac{\ddot{a}}{a}\,, \quad R_{ij} = -\delta_{ij}\frac{2K + 2\dot{a}^2 + a\ddot{a}}{(1 + K\mathbf{x}^2/4)^2}\,, \quad R = -6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{K}{a^2}\right]\,, \tag{1.42}$$

Since from this we derive $G_{00} = -3H^2$, the 00-component of the EE's in (1.22) is the easily derived

$$3M_{\rm Pl}^2\left(H^2 + \frac{K}{a^2}\right) = \sum_i \rho_i\,, \tag{1.43}$$

where i runs over all constituents of the universe, such as for example photons, dark matter, neutrinos or baryons. This is known as the *Friedmann equation*. Notice that since an FLRW metric is specified by a single function $a(t)$, we need only one of the ten EEs to determine the solution.
— It is often convenient to use the dimensionless *fractional energy densities*[4] Ω_i we defined early, as opposed to using the dimension-ful ρ_a. Recall that

$$\rho_{crit}(t) \equiv 3M_{\rm Pl}^2 H^2(t) \quad \Rightarrow \quad \Omega_a \equiv \frac{\rho_a(t)}{3H^2(t)M_{\rm Pl}^2}\,. \tag{1.44}$$

The fractional energy densities at the present time ($a = a_0$), indicated by a subscript 0, are worth remembering:

$$\Omega_{\Lambda,0} = 0.72\,, \quad \Omega_{b,0} = 0.04\,, \quad \Omega_{DM,0} \simeq 0.24\,, \quad \text{and} \quad \Omega_{r,0} = 3.4\times 10^{-5}\,. \tag{1.45}$$

[4] Beware of different conventions. Sometimes $\Omega_a(t)$ is defined in terms of the time dependent critical energy density $3M_{\rm Pl}^2H^2$ and/or the time dependent density $\rho(a)$ and sometimes it is just its value today, which we indicate as $\Omega_{a,0}$ to avoid confusion.

Here $\Omega_{\Lambda,0}$ stands for the cosmological constant, $\Omega_{b,0}$ for baryons, $\Omega_{DM,0}$ for dark matter and $\Omega_{r,0}$ for radiation (photons and relativistic neutrinos). The time evolution is then simply given by

$$\rho_i(t) = 3M_{\rm Pl}^2 H_0^2 \frac{\Omega_{i,0}}{a(t)^{3(1+w)}} \,. \tag{1.46}$$

This gives a simple expression for the Hubble parameter as function of the scale factor:

$$H^2 = H_0^2 \sum_i \frac{\Omega_{i,0}}{a(t)^{3(1+w)}} \,. \tag{1.47}$$

Note that it is customary to express the fraction of density $\Omega_{i,0}$ multiplied by h^2 defined by

$$H_0 = 100 \times h \; \frac{\rm km}{\rm sec\, Mpc} \,. \tag{1.48}$$

By using $\Omega_{i,0}h^2$ instead of $\Omega_{i,0}$ one is immune to changes or errors in the measurement of H_0. In other words, measurements of the actual energy density $\rho_{i,0}$ can be converted into $\Omega_{i,0}h^2$ without assuming the value of the Hubble parameter. This is particularly important because since 2018 local and cosmological measurement of H_0 present a 3-4 σ tension. In particular CMB/BAO measurements give $h = 67.6 \pm 0.6$ [1] while local measurements based on the distance ladder, which find $h = 73.24 \pm 1.74$ [53]. In this way the measurement of ρ_m is not contaminated by the error on H_0. Notice that $h^{-2} \simeq 2$.
—- It is then convenient to divide each side of the equation by the so-called *critical density*, which in our conventions is a function of time,

$$\rho_c \equiv 3M_{\rm Pl}^2 H^2 \,. \tag{1.49}$$

This is the total energy density that for a given value of H corresponds to a spatially-flat universe. Dividing the Friedmann equation by ρ_c we find

$$1 - \Omega_k = \sum_a \Omega_a \,, \quad \text{with} \quad \Omega_k \equiv -\frac{K}{H^2a^2} \,, \quad \Omega_a \equiv \frac{\rho_a}{\rho_c} \,. \tag{1.50}$$

Here we introduced the dimensionless fractional energy densities Ω_a. Notice that Ω_k can be negative for closed universes with $K > 0$.

The second Friedmann equation and the acceleration equation There are two other combinations of EE's that come in handy. First, by taking the time derivative of the Friedmann equation and using the continuity equation to get rid of $\dot\rho$, we can find the variation of the Hubble parameter

$$\dot H - \frac{K}{a^2} = -\frac{1}{2M_{\rm Pl}^2} \left(\rho + p\right) \,. \tag{1.51}$$

This equation is sometimes known as the *second Friedmann equation*. Most cosmological "stuff" obeys the Null Energy Condition (1.57), and so in a flat universe H decreases during cosmic evolution.

Noticing that $\ddot{a}/a = \dot{H} + H$ and using the two Friedmann equation one finds the *acceleration equation*

$$\frac{\ddot{a}}{a} = -\frac{1}{6M_{\rm Pl}^2}\left(\rho + 3p\right) \,. \tag{1.52}$$

Note that the curvature term cancelled out. This equation is useful because it gives us the condition on the matter sector under which the universe accelerates or decelerates.

Solutions Using a to parameterize time and solving for $H(a)$ algebraically and hence for $a(t)$ finding

$$H = \frac{\dot{a}}{a} = \sqrt{\frac{\rho}{3M_{\rm Pl}^2}} = H_0 \left(\frac{1}{a}\right)^{3(1+w)/2} \,. \tag{1.53}$$

This first order differential equation can be integrated to give

$$\int_{t_0}^{t} dt' = \int_1^a \frac{d\tilde{a}}{\tilde{a}H_0(\tilde{a})^{-3(1+w)/2}} \,, \tag{1.54}$$

where we used the convention $a(t_0) = 1$. For $w > -1$, the solution $a(t)$ has the property that $a(t_{\rm BB}) = 0$ for some time $t_{\rm BB}$. This moment in time is the big bang. If we choose to start counting time from this moment, i.e. we shift time such that $t_{\rm BB} = 0$, we find

$$a(t) = \left[\frac{3}{2}\left(1+w\right)H_0 t\right]^{\frac{2}{3(1+w)}} = \left(\frac{t}{t_0}\right)^{\frac{2}{3(1+w)}} \,, \tag{1.55}$$

where we recognized that

$$t_0 = \frac{2}{3(1+w)}\frac{1}{H_0}$$

is the age of the universe, i.e. the time elapsed since the big bang until today. Note that this is a fully non-linear solution of Einstein's equations. It is quite remarkable that it happens to describe our universe so well on large scales. This solution is simple because we have assumed that all the content of the universe obeys the same equation of state with the same constant w. For a more realistic universe containing different substances one has to solve this equation numerically. It is sometime useful to find the scale factor in terms of conformal time, $a(\tau)d\tau = dt$. For a single fluid universe this is given by

$$a(\tau) = \left(\frac{\tau}{\tau_0}\right)^{\frac{2}{3w+1}} \,. \tag{1.56}$$

Important solutions for the scale factor are found in each of the cases below:

- For non-relativistic matter, or *dust*, $w \simeq 0$ so

$$a = \left(\frac{t}{t_0}\right)^{2/3} = \left(\frac{\tau}{\tau_0}\right)^2 \,.$$

Box 1.1 Null Energy Condition (NEC) A certain form of matter with energy-momentum tensor $T_{\mu\nu}$ satisfies the Null Energy Condition if for every null vector $N^\mu N_\mu = 0$ one has

$$T_{\mu\nu} N^\mu N^\nu \geq 0 \quad \text{(NEC)}\,. \tag{1.57}$$

Using the perfect fluid parameterization in (1.34), this implies $\rho + p \geq 0$. Violations of the NEC are often associated with pathologies such as ghosts instabilities (i.e. field with the wrong-sign kinetic term that can be nucleated by decreasing the energy of the system) or tachyon instabilities [25]. Yet, more exotic theories with non-standard kinetic terms, such as the ghost condensate, are known to safely violate the NEC, see e.g. [19, 54].

- For relativistic matter, or *radiation*, $w = 1/3$ so
$$a = \left(\frac{t}{t_0}\right)^{1/2} = \left(\frac{\tau}{\tau_0}\right)\,.$$
- For a cosmological constant, or *vacuum energy*, $w = -1$ this expressions is singular. In this case one has to go back and solve the Friedmann equation again in the presence of a cosmological constant,
$$H^2 = \text{const} = H_0^2 \quad \Rightarrow \quad a(t) = e^{H_0(t-t_0)} = -\frac{1}{H\tau}\,.$$
Note that in this case the big bang has been pushed all the way to $t_{\rm BB} = -\infty$ and that $-\infty < \tau < 0$.

1.4 Symmetric spaces*

In this subsection, which is not examinable, we discuss how to describe symmetries in GR and derive properties of maximally symmetric spaces.
Some common solutions of GR display symmetries, exact or approximated. A spacetime enjoys an *isometry* if there is a change of variables $x \to \tilde{x} = \tilde{x}(x)$ that leaves the metric unchanged in the sense that

$$\tilde{g}_{\mu\nu}(\tilde{x}) = \frac{\partial x^\rho}{\partial \tilde{x}^\mu}\frac{\partial x^\sigma}{\partial \tilde{x}^\nu} g_{\rho\sigma}(x) \stackrel{!}{=} g_{\mu\nu}(\tilde{x})\,. \tag{1.58}$$

This is equivalent to (prove it)

$$ds^2(x) = g_{\mu\nu}(x) dx^\mu dx^\nu \stackrel{!}{=} g_{\mu\nu}(\tilde{x}) d\tilde{x}^\mu d\tilde{x}^\nu = ds^2(\tilde{x}(x))\,, \tag{1.59}$$

where by $ds^2(\tilde{x}(x))$ means that we substituted every x with an $\tilde{x}(x)$. Isometries are best discussed using Killing vectors. Given the change of coordinates $x'^\mu = x^\mu + \xi^\mu$, every tensor changes by minus its Lie derivative $\mathcal{L}_\xi$ (see Box A) in the ξ direction. For the metric to be invariant we require[5]

$$\Delta g_{\mu\nu}(x) \equiv g'_{\mu\nu}(x) - g_{\mu\nu}(x) = -\mathcal{L}_\xi g_{\mu\nu}(x) \tag{1.60}$$
$$= -\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu \stackrel{!}{=} 0\,. \tag{1.61}$$

[5] CFU: *Derive this expression taking advantage of the fact that the metric is covariantly constant, $\nabla_\mu g_{\nu\rho} = 0$, and using Eq. (1.18).*

Vectors ξ for which $\mathcal{L}_\xi g_{\mu\nu} = 0$ leave the metric invariant and are called Killing vector fields, or simply *Killing vectors*. Remarkably, Killing vectors are completely determined by their value and that of their derivative at one point. To see this, recall two defining properties of the Riemann tensor

$$[\nabla_\mu, \nabla_\nu] V_\rho = R_{\rho\sigma\mu\nu} V^\sigma \,, \quad R_{\rho\sigma\mu\nu} + R_{\rho\nu\sigma\mu} + R_{\rho\mu\nu\sigma} = 0\,. \tag{1.62}$$

We can sum over cyclic permutations of the first equation, use the second equation as well as the definition of Killing Vectors (1.61) and find

$$\nabla_\rho \nabla_\sigma \xi_\mu = [\nabla_\rho, \nabla_\mu] \xi_\sigma = R_{\lambda\sigma\mu\rho} \xi^\lambda\,. \tag{1.63}$$

The solution of this second order pde are determined by the initial condition $\{\xi^\mu(\bar{x}), \nabla_\nu \xi^\mu(\bar{x})\}$ at some point $\bar{x}$ (at least locally one can construct the solution as a Taylor expansion in $x - \bar{x}$), and so takes the form

$$\xi_\mu(x) = A^\rho_\mu(x, \bar{x}) \xi_\rho(\bar{x}) + B^{\sigma\rho}_\mu(x, \bar{x}) \nabla_\sigma \xi_\rho(\bar{x})\,. \tag{1.64}$$

Since we can specify at most D independent $\{\xi^\mu(\bar{x})\}$ and $D(D-1)/2$ independent $\{\nabla_\nu \xi^\mu(\bar{x})\}$ (antisymmetry of $\nabla_\mu \xi_\nu$ follows from the Killing condition), the maximum number of a isometries a spacetime can enjoy is $D(D+1)/2$, which reduced to 10 in $D = 4$. Spaces that saturate this upper bound on the number of independent Killing vectors (isometries) are referred to as *maximally symmetric spaces*[6].
To gain some intuition on these $D(D+1)/2$ generators, let us choose the following $D(D+1)/2$ linearly independent initial conditions

$$\begin{cases} \xi^{(\alpha)}_\rho = \delta^\alpha_\rho\,, \\ \nabla_\rho \xi^{(\alpha)}_\sigma = 0\,. \end{cases} \qquad \text{(D solutions)}$$
$$\begin{cases} \xi^{(\alpha\beta)}_\rho = 0\,, \\ \nabla_\rho \xi^{(\alpha\beta)}_\sigma = \delta^\alpha_{(\rho} \delta^\beta_{\sigma)}\,. \end{cases} \qquad (D(D-1)/2 \text{ solutions})\,, \tag{1.65}$$

where the indices α and β label the solutions. The first D solutions cover completely the tangent space of the manifold at point $\bar{x}$ and can hence be thought of as generalised[7] translations: they move the (arbitrary) point $\bar{x}$ in any of the D direction. It can be proven (see Ch 13 of [64]) that the remaining $D(D-1)/2$ Killing vectors change any vector $V^\mu(\bar{x})$ into any other vector $\tilde{V}^\mu(\bar{x})$ with the same norm $V^\mu V_\mu = \tilde{V}^\mu \tilde{V}_\mu$. These isometries can then be thought of as generalised[8] rotation. We conclude that a maximally symmetric space is *homogeneous* (invariant under generalised translations) and *isotropic* (invariant under generalised rotations)[9]. The converse

[6]CFU: *If we can freely specify the initial conditions $\{\xi^\mu(\bar{x}), \nabla_\nu \xi^\mu(\bar{x})\}$, why are not all spaces maximally symmetric? There are "integrability" restrictions, which depend on the metric, on the set of initial data $\{\xi^\mu(\bar{x}), \nabla_\nu \xi^\mu(\bar{x})\}$ that admits a solution (see 13.1.12 of [64]).*

[7]"Generalised" here means both that these act as translations only locally, rather than globally, and that it is not specified whether the coordinates they translate are Euclidean or not.

[8]"Generalised" here also refers to the fact that when the signature of the metric is e.g. Lorentzian, rather than Minkowskian, these isometries correspond locally to boosts rather than rotations.

[9]CFU: *Does homogeneity imply isotropy? (no, e.g. this room with constant gravitational field) Does isotropy imply homogeneity? (no around a single point) Does isotropy around every point imply homogeneity? (yes)*

is also true, i.e. all homogeneous and isotropic spacetimes are maximally symmetric as follows from a simple counting of isometry generators[10].
There are three more theorems that we will quote without proof:

- *Uniqueness:* Maximally symmetric spaces are uniquely characterised by the value of the Ricci scalar R, which is just a constant number over the space by homogeneity, and the signature of the metric (see 13.2 of [64]).
- For Maximally symmetric spaces the Riemann tensor is proportional to the metric

$$R_{\mu\nu\rho\sigma} = K g_{\mu(\sigma} g_{\nu\rho)} \,, \tag{1.66}$$

where K is related to the Ricci scalar by

$$R = -D(D-1)\, K \,. \tag{1.67}$$

- If a space M contains a maximally symmetric subspace $N \subset M$, the metric can always be written as the following "warp product"

$$ds^2 = g_{ab}(x)dx^a dx^b + f(x)\tilde{g}_{ij}(y)dy^i dy^j \,, \tag{1.68}$$

with $\tilde{g}_{ij}$ the metric of the maximally symmetric subspace, y the coordinates of the subspace and x the remaining coordinates.

Problems for lesson 2

P.1.1 Compute the evolution of the entropy density $s(a)$ in an FLWR universe in the (good) approximation that all processes are adiabatic. How does this compare to the evolution of any other conserved charge?

P.1.2 Compute and solve the geodesic equation for a massless particle in FLRW

P.1.3 Using the definition of isometry for a coordinate transformation $\tilde{x}(x)$, namely

$$\tilde{g}_{\mu\nu}(\tilde{x}) = g_{\mu\nu}(\tilde{x}) \,, \tag{1.69}$$

verify that the (flat) FLRW metric

$$ds^2 = -dt^2 + a^2 dx^i \delta_{ij} dx^j \tag{1.70}$$

is indeed homogeneous and isotropic (i.e. isometric with respect to spatial translations $\tilde{x}^i = x^i + b^i$ and rotations $\tilde{x}^i = R^i_j x^j$).

P.1.4 Solve the Friedmann equation for $w = -1$ and $w \neq -1$.

P.1.5 Consider an FLRW universe with a single fluid and derive the acceleration equation Eq. (1.52) for $\ddot{a}(t)$. You can either derive the ii component of the Einstein equations or use the Friedman equation together with the covariant conservation of energy, $T^{0\mu}_{;\mu} = 0$. For what w does one get accelerated expansion?

[10] CFU: *Prove this. Prove also that the number of Killing vectors does not depend on the choice of coordinates.*

2 Redshift and distances

In this chapter we introduce the key cosmological concepts of redshift and distance, which provide the fundamental link between theoretical models and astronomical observations. We begin with the notion of cosmological redshift, arising from the stretching of photon wavelengths in an expanding universe, and contrast it with the familiar Doppler effect. Building on this, we define several operationally distinct but observationally crucial notions of distance, including comoving, luminosity, and angular diameter distances, as well as the particle horizon and the age of the universe. These quantities are then connected to the composition of the universe—curvature, radiation, baryons, neutrinos, dark matter, and dark energy—laying the groundwork for interpreting observational probes of cosmic expansion and structure formation.

2.1 Geodesics and redshift

Olbers' paradox is the argument that the universe cannot be eternal and infinite because otherwise the night sky should be bright, since every direction in the sky would point to some star with a similar intrinsic luminosity as the sum. In the Big Bang model, the universe has a finite age (about 13.7 billion years). This actually makes the problem worse since the universe must have been much hotter in the past and we should see an even brighter sky from very hot thermal radiation. The resolution is that in FLRW the wavelength of light redshifts $E \sim \lambda^{-1} \sim a^{-1}$. To discuss this we start with a short review of geodesics in general relativity.

The geodesic equation In general relativity, the four-velocity of a particle moving along a trajectory $X^\mu(s)$ is defined as $U^\mu(s) \equiv dX^\mu/ds$, where s is the proper time. Particles follow geodesics, which extremize the proper time along their worldline. Extremizing the relativistic action $-m\int_A^B ds$ leads to the geodesic equation

$$\frac{dU^\mu}{ds} + \Gamma^\mu_{\alpha\beta} U^\alpha U^\beta = 0. \tag{2.1}$$

Using the chain rule, one can write

$$\frac{dU^\mu}{ds} = \frac{dX^\alpha}{ds}\frac{\partial U^\mu}{\partial X^\alpha} = U^\alpha \partial_\alpha U^\mu.$$

For massive particles it is convenient to define the four-momentum $P^\mu = mU^\mu$, so that the geodesic equation becomes

$$P^\alpha \partial_\alpha P^\mu + \Gamma^\mu_{\alpha\beta} P^\alpha P^\beta = 0. \tag{2.2}$$

Although derived for massive particles, the same form holds for photons in the massless limit. We now specialize to the FRW metric, written as

$$ds^2 = -dt^2 + a^2(t)\gamma_{ij} dx^i dx^j.$$

In this case the only non-vanishing Christoffel symbols are (1.21)

$$\Gamma^0_{ij} = a\dot{a}\,\gamma_{ij}, \qquad \Gamma^i_{0j} = \frac{\dot{a}}{a}\delta^i_j, \qquad \Gamma^i_{jk} = \tfrac{1}{2}\gamma^{il}(\partial_j\gamma_{kl} + \partial_k\gamma_{jl} - \partial_l\gamma_{jk}), \tag{2.3}$$

together with those related by symmetry. Homogeneity implies $\partial_i P^\mu = 0$, so the geodesic equation reduces to

$$P^0 \frac{dP^\mu}{dt} = -\Gamma^\mu_{\alpha\beta} P^\alpha P^\beta = -(2\Gamma^\mu_{0j} P^0 + \Gamma^\mu_{ij} P^i) P^j. \tag{2.4}$$

For a particle initially at rest ($P^j = 0$), this implies $dP^\mu/dt = 0$, i.e. the particle remains at rest.

Redshift and Hubble's law. Let us now consider the $\mu = 0$ component of the above relation. Setting $P^0 = E$,

$$E\frac{dE}{dt} = -\Gamma^0_{ij} P^i P^j = -\frac{\dot{a}}{a} p^2, \tag{2.5}$$

where the physical three-momentum p is defined as

$$p^2 \equiv g_{ij} P^i P^j = a^2 \gamma_{ij} P^i P^j. \tag{2.6}$$

Using $E^2 - p^2 = m^2$ it follows that $EdE = pdP$, and so from (2.5) one finds

$$\frac{\dot{p}}{p} = -\frac{\dot{a}}{a} \qquad \Rightarrow \qquad p \propto \frac{1}{a}. \tag{2.7}$$

Hence the physical momentum of any particle decreases as the universe expands. For massless particles, $E = p \propto 1/a$, so that

$$\lambda_0 = \frac{a(t_0)}{a(t_1)} \lambda_1, \tag{2.8}$$

i.e. wavelengths stretch in proportion to the scale factor, a phenomenon known as *cosmological redshift*. An alternative derivation using the conformal period of classical waves is presented in Box **??** and leads to the same relation.

The redshift z is defined by

$$z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1}, \tag{2.9}$$

which directly implies

$$1 + z = \frac{1}{a(t_1)}. \tag{2.10}$$

For nearby sources, expanding $a(t_1)$ gives

$$a(t_1) = a(t_0)\left[1 + (t_1 - t_0)H_0 + \cdots\right], \tag{2.11}$$

where H_0 is the Hubble constant. Thus $z \simeq H_0(t_0 - t_1)$. Since for nearby objects $t_0 - t_1 \simeq d$, one recovers Hubble's law

$$z \simeq H_0 d + \cdots, \tag{2.12}$$

showing that redshift increases linearly with distance. This is precisely the relation first discovered by Hubble, demonstrating that redshift-distance measurements imply $\dot{a} > 0$, i.e. that the universe is expanding. A modern Hubble diagram from type Ia supernovae is shown in Fig. 4. It is customary to write $H_0 = 100\,h\,\mathrm{km/s/Mpc}$, introducing the dimensionless parameter h. Current Planck data gives $h = 0.6766 \pm 0.0042$, though local measurements disagree at the few percent level, leading to the so-called Hubble tension. The observation of Hubble's law therefore provides direct evidence for an expanding universe and motivates the study of cosmic dynamics.

Alternative derivation An alternative derivation of cosmological redshift proceeds as follows. Consider now the light propagating to us along the $-\hat{r}$ direction from some emitting source at comoving position $\{r, \theta, \phi\}$ in spherical coordinates. Photons are massless and so follow null geodesics with null tangent vector

$$ds^2 = 0 \quad \Rightarrow \quad \frac{dt}{a} = dr\,. \tag{2.13}$$

Consider a wave crest being emitted at some time t_e and arriving at time t_o to the observer at the origin $r = 0$. Then

$$\int_{t_e}^{t_o} \frac{dt}{a} = \int_0^r dr = r\,. \tag{2.14}$$

Consider now the subsequent wave crest being emitted at some time $t_e + \lambda_e$ (recall our units $c = 1$) and arriving at time $t_o + \lambda_o$ to the observer at the origin $r = 0$. We have similarly

$$\int_{t_e+\lambda_e}^{t_o+\lambda_o} \frac{dt}{a} = \int dr = r\,. \tag{2.15}$$

Subtract (2.14) from (2.15) and find

$$\int_{t_e}^{t_e+\lambda_e/c} \frac{dt}{a} = \int_{t_o}^{t_o+\lambda_o/c} \frac{dt}{a} \tag{2.16}$$

Performing the integral on each side under the approximation that a doesn't change much[11] we find

$$\frac{\lambda_e}{a_e} = \frac{\lambda_o}{a_0} \Rightarrow \frac{\lambda_o}{\lambda_e} = \boxed{1 + z = \frac{1}{a_e}}\,, \tag{2.17}$$

where we used the fact that in all practical application the observer is us today and so $a_0 = 1$ by convention.

To conclude we note that cosmological redshift is not Doppler redshift. The two agree *only at linear order*, i.e. $v \simeq zc + \mathcal{O}(z^2)$, which is valid on distances much smaller than the spacetime curvature H^{-1} of FLRW. This relation follows simply from dimensional analysis and the equivalence principle. To see this requires the definition of comoving distance Eq. (2.20) discussed below.

2.2 Distances

In non-relativistic mechanics there are many different ways to measure distance, such as using a ruler, observing the apparent size of an object of known intrinsic size, observing the luminosity of a known "standard" candle and so on. In general relativity, different measurements give different answers, so we have many different concepts of "distance" depending on how it is determined operationally. A few types of distances are used routinely in cosmology, for example to probe the expansion history of the universe. All physical distances are conveniently related to comoving coordinates with appropriate factors of a.

[11] CFU: *Estimate the change in $a(t)$ during one period of visible light*

Comoving distance For an FLRW metric, the *comoving distance* is the distance travelled by a photon in a certain time interval in comoving coordinates. Consider spherical comoving coordinates

$$ds^2 = a^2\left[-d\tau^2 + d\chi^2 + \chi^2\left(d\theta^2 + \sin\theta^2 d\phi^2\right)\right] , \tag{2.18}$$

and recall that for null geodesics $ds^2 = a^2\left[-d\tau^2 + d\eta^2\right] = 0$. Choosing coordinates such that the propagation takes place exclusively in the χ direction we find $d\tau = d\eta$ and $\tau = \eta + \text{const.}$ For the comoving distance one finds

$$\chi(t_i, t_f) = \int d\tau = \int_{t_i}^{t_f} \frac{dt}{a(t)} = \int \frac{da}{a^2 H(a)} = \int \frac{dz}{H(z)} . \tag{2.19}$$

A common case is when the photon arrive on Earth now and so $t_f = t_0$ or $a(t_f) = a_0 = 1$. Then one finds the comoving distance to a given redshift

$$\chi(z) = \int_0^z \frac{dz}{H(z)} . \tag{2.20}$$

For nearby points at redshift $z \ll 1$ we can estimate χ by

$$\chi(z) \simeq \frac{1}{H_0}\left[z + \frac{1 - \ddot{a}_0/H_0^2}{2} z^2 + \dots\right] ,$$

where $\ddot{a}_0$ is the current acceleration.

Particle horizon Let us define the *particle horizon* $d_{\text{p.h.}}$ as the maximum physical distance that light can have traveled since some "beginning of time" t_i, which could also be infinite. Any place further than that, at distance $d > d_{\text{p.h.}}$ cannot have sent us any signal. We are not inside their future light cone, they are not in our past light cone. The particle horizon is then given by [Problem P.2.1]

$$d_{\text{p.h.}}(t) \equiv a(t)\chi(t_i, t) = a(t)\tau(t_i, t) = a(t)\int_{t_i}^{t} \frac{dt'}{a(t')} , \tag{2.21}$$

where χ is the comoving distance and $a(t)$ transforms it into a physical distance. To gain intuition let us consider some simple single component universe, for which the scale factor is a power law in time (1.53), $a \propto t^{2/(3+3w)}$, with $w > -1$ for expansion (the case $w = -1$ is straightforward but requires a separate discussion) and beginning of time $t_i = 0$. Then

$$d_{\text{p.h.}}(t) = t^{2/(3+3w)} \int_0^t \frac{dt'}{t'^{2/(3+3w)}} = \frac{t^{2/(3+3w)}}{1 - 2/(3+3w)}\left[t^{1-2/(3+3w)} - 0^{1-2/(3+3w)}\right] . \tag{2.22}$$

For $2/(3+3w) > 1$, or equivalently $w < -1/3$ this diverges, while it converges to $d_{\text{p.h.}} \propto t$ (as expected by dimensional analysis) for $w > -1/3$. For example, $d_{\text{p.h.}} = 3t$ for matter ($w = 0$) and $d_{\text{p.h.}} = 2t$ for radiation ($w = 1/3$).

Luminosity distance The *luminosity distance* is useful as a measurement of the expansion of the universe when observing an object of known luminosity. The intrinsic luminosity L, or simply luminosity, is the total amount of energy radiated per unit time. In Euclidean geometry, the intrinsic luminosity is related to the observed flux f by

$$f \equiv \frac{L}{4\pi d_L^2} \,, \tag{2.23}$$

where d_L is the luminosity distance and the factor of 4π comes about because f is defined as observed energy per unit time per unit surface and we assume the object emits light isotropically. In an expanding universe, we choose to *define* d_L via the same relation (2.23). However, we now have to account for three factors:

- The comoving distance from emitting source and observation is $\chi(t_e, t_o)$ and it gets a factor of a_o to become a physical distance.
- The rate of arrival of photons decreases by a factor of a_e/a_o. Assuming $a_o = a_0 = 1$ this becomes $a_e/a_o = a_e = (1+z)^{-1}$.
- The energy of incoming photons is redshifted by another factor of $a_e/a_o = (1+z)^{-1}$.

Putting things together

$$f = \frac{L}{4\pi\,(\chi a_o)^2}\left(\frac{a_e}{a_o}\right)^2 \equiv \frac{L}{4\pi d_L^2} \,, \tag{2.24}$$

where in last step we assumed z is the redshift of emission and that observation takes place today at $a_o = a_0 = 1$. Hence we find

$$d_L(z) = \chi\left(\frac{1}{a_e}\right) = (1+z)\,\chi(z) \,. \tag{2.25}$$

Instead of luminosity and flux, astronomers prefer to use two related quantities, whose origin dates back to the seventh century AD. Ptolemy made a survey of stars visible to the naked eye and divided them into six groups, from brightest in group one to faintest in group six. In 1856, Nortman Pogson proposed a mathematically precise classification that extended Ptolemy's heuristic approach. He introduced the *apparent magnitude* m as

$$m \equiv -2.5 \log\left(\frac{l}{l_0}\right) \,, \tag{2.26}$$

where $l_0 = 2.5 \times 10^5$ erg cm^{-2} s^{-1} is a reference flux. Similarly one introduces the *magnitude* M by

$$M \equiv -2.5 \log\left(\frac{L}{L_0}\right) \,, \tag{2.27}$$

where $L_0 = 3\times 10^{35}$ erg s^{-1} is a reference luminosity. For a reference, $m_{\rm sun} = -27$, $m_{\rm Sirius} \simeq -1$, $m_{\rm M31} \simeq 0$ and the faintest object we can see by eye has $m \sim 6$. Modern telescopes can see much further. For example, the limit of the Hubble space telescope using visible light is

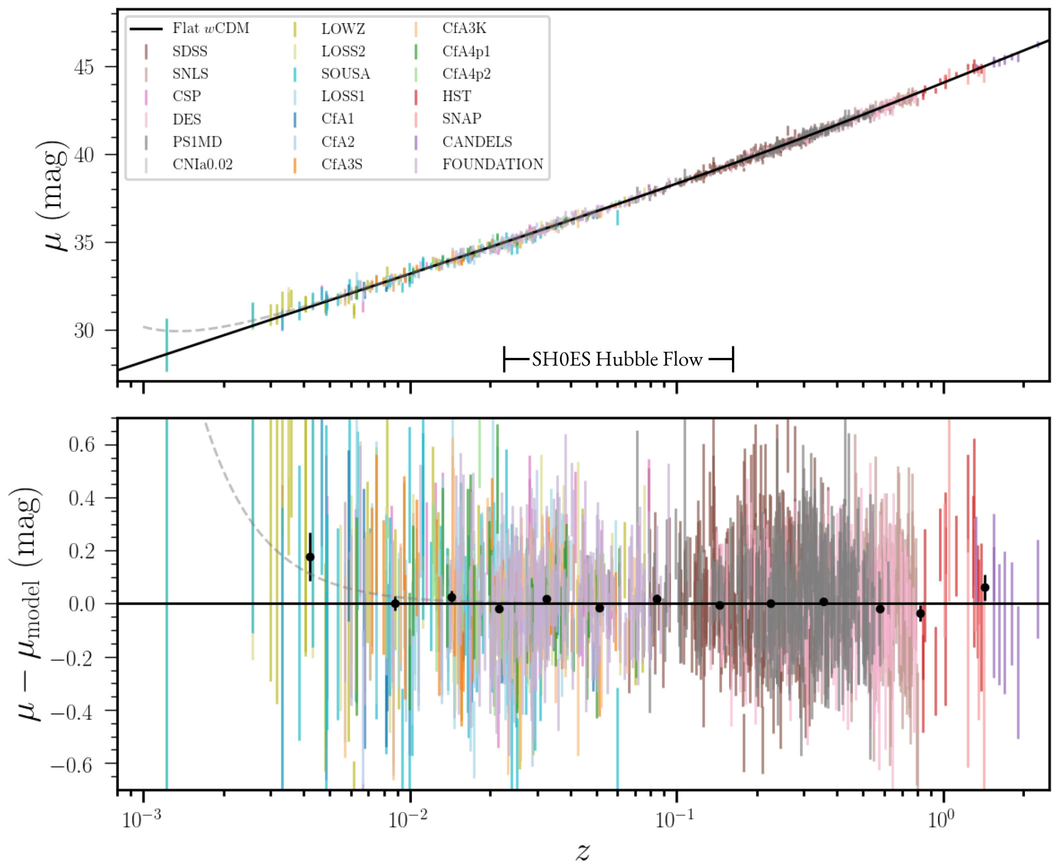


Figure 4: The Hubble diagram for the Pantheon+ catalog showing the modulus distance μ as function of redshift for about 1500 type Ia supernovae. Figure from [10].

approximately $m \leq 30$. When working with magnitudes, the quantity that is sensitive directly to the luminosity distance d_L is the *distance modulus*

$$\mu \equiv m - M = 5 \log_{10} \left(\frac{d_L}{10\text{pc}} \right) . \tag{2.28}$$

The distance modulus μ can be plotted as function of redshift z to obtain the general relativistic extension of the Hubble diagram. This has been done mainly with supernovae of type Ia, which are believed to be standardizable candles for which the absolute magnitude can be known. The result for the Pantheon plus supernovae catalog is shown in Fig. 4.

Angular diameter distance. The *angular diameter distance* in Euclidean geometry is defined for objects of physical size r_{ph} subtending and angle θ from the point of view of the observer by $r_{\text{ph}} \equiv d_A \theta$. Here there is only one factor of a in the relation of comoving to physical distance, hence

$$d_A(z) = \frac{\chi(z)}{1+z} . \tag{2.29}$$

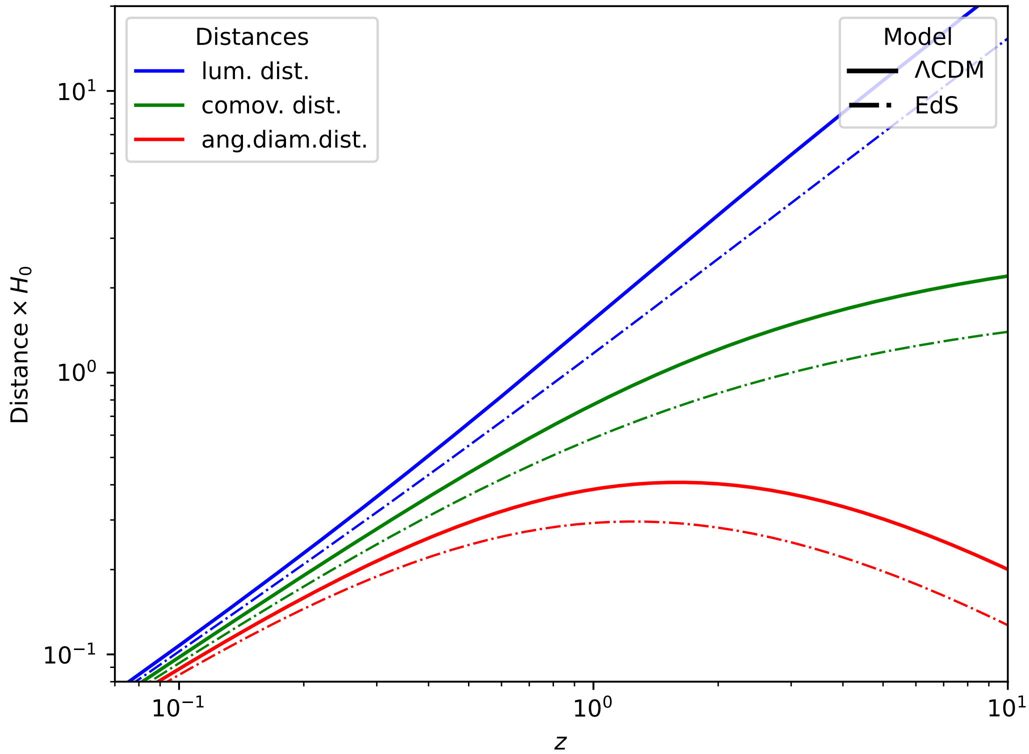


Figure 5: Angular diameter, comoving and luminosity distances as function of redshift. Continuous lines represent ΛCDM, namely a flat FLRW universe with $\Omega_\Lambda = 0.7$ and $\Omega_m = 0.3$. The dot-dashed lines refer to a flat FLRW universe with $\Omega_\Lambda = 0$ and $\Omega_m = 1$, a.k.a. an Einstein de Sitter (EdS) universe.

The angular diameter distance is difficult to use in practice because objects such as supernovae or galaxies at cosmological distances do not have well defined edges from which to extract a size. However, two important exception are Baryon Acoustic Oscillations (BAO) and acoustic oscillations in the CMB. These observables possess features in the distribution of galaxies and photons respectively, which were imprinted by plasma oscillations before the universe became transparent. When discussing the CMB for example, the characteristic scale is the sound horizon at recombination r_s. This scale is often described by its comoving distance r_s rather than physical distance. As a result, the apparent angular scale θ_* in the sky can be written in the following two ways

$$\theta_* = \frac{r_{\rm ph}}{d_A} = \frac{r_s}{\chi}\,,$$

where $r_{\rm ph}$ and d_A are physical distances and r_s and χ are comoving.
The various distances are plotted in Fig. 5 for two different flat cosmologies, one with a cosmological constant and one without. This figure makes it clear that measuring these distances gives us a way to distinguish different models of the content of the universe. Also, it is clear that different ways of measuring distance in an expanding universe give different values, but they all agree as $z \to 0$ since that's when there has not been enough time for the expansion of

the universe to have a sizable effect. Notice that the angular diameter distance in our universe starts decreasing around $z \sim 1$. This is analogous to what happens to d_A for a fixed-size $r_{\rm ph}$ segment on the surface of a sphere. As the segment recedes from the observer at the north pole, its angular size θ decreases at first but then starts increasing again after the segment crosses the equator. The angular diameter distance $d_A = r_{\rm ph}/\theta$ has the opposite behaviour.

Age of the universe The *age of the universe* is computed from

$$t_{\rm age} = \int dt = \int \frac{da}{\dot{a}} = \int \frac{da}{aH(a)} = \int \frac{da}{a} \left[\frac{\sum_i \rho_i}{3M_{\rm Pl}^2}\right]^{-1/2} , \tag{2.30}$$

where $H(a)$ is derived from the Friedmann equation. Notice that one does not need $a(t)$ here. To make progress we have to state the constituents of our universe. As we will see in the next section, the most successful cosmological model, which is known as Λ Cold Dark Matter (LCDM), contains non-relativistic matter $\Omega_{m,0} = \Omega_{b,0} + \Omega_{DM,0}$, relativistic matter Ω_r from photons and neutrinos and a cosmological constant. For LCDM, using the definitions in (1.50) and the solution of the continuity equation in (1.50), the age of the universe is

$$t_{\rm age} = \frac{1}{H_0} \int_0^1 \frac{da}{a} \left[\Omega_{\Lambda,0} + \Omega_{m,0} a^{-3} + \Omega_{r,0} a^{-4}\right]^{-1/2} , \tag{2.31}$$

This can be readily evaluated numerically to give $t_{\rm age} \simeq 0.95 H_0^{-1} \simeq 13.7\,{\rm Gyr}$ for the current best fit values of the cosmological parameters, approximately

$$\Omega_{\Lambda,0} \simeq 0.7\,, \qquad \Omega_{m,0} \simeq 0.3\,, \qquad \Omega_{r,0} \simeq 9 \times 10^{-5}\,. \tag{2.32}$$

Problems for lesson 2

P.2.1 Compute the particle horizon for a matter and radiation dominated universe of fixed age. Which one is larger? Interpret your answer.

P.2.2 (Mukhanov's book ex 1.9) By embedding a 3D sphere (pseudo-sphere) in a $(3+1)$D Euclidean (Lorentzian) space, verify that the metric of a 3D space of constant curvature can be written as

$$dl^2 = R^2 \left[\frac{dr^2}{1-kr^2} + r^2 \left(d\theta^2 + \sin\theta^2 d\phi^2\right)\right] , \tag{2.33}$$

where $R > 0$ and $k = 0, \pm 1$.

3 Constituents of the Universe

We will now review the five main components of the Universe: photons, baryons, neutrinos, dark matter and dark energy.
Only particles with a lifetime comparable with the age of the Universe have a sizable density today. Within the standard model of particle physics, we have photons, protons, electrons and neutrinos. Free neutrons decay in about 15 minutes, but they can be stable when combined with protons to form the nuclei of atoms.

3.1 Photons

The density of *photons* can be derived straightforwardly from the temperature of the Cosmic Microwave Background (CMB), $T_{\rm CMB} = 2.72548 \pm 0.00057$ K [27, 28]. We know that the dimensionless chemical potential, which is defined in Eq. (6.9), is small[12] $\mu < 6 \times 10^{-5}$, so we can use the exact Planck black body spectrum as in (6.22):

$$\rho_\gamma = 3 p_\gamma = \frac{\pi^2}{15} T_{\rm CMB}^4 \,, \tag{3.1}$$

where we used that a photon is a boson with two degrees of freedom $g_\gamma = 2$ (helicities ± 1). From the covariant conservation of energy we know that $\rho_\gamma \propto a^{-4}$ and therefore for photons $T\,a = \text{const}$. Finally, one finds

$$\Omega_\gamma h^2 = 2.5 \times 10^{-5} \quad \rightarrow \quad \Omega_\gamma \simeq 5 \times 10^{-5} \,. \tag{3.2}$$

3.2 Baryons

In particle physics, the word *baryons* strictly speaking refers to protons and neutrons but in cosmological lingo it is customary to include electrons as well. The Universe appears to be neutral as a whole, so we will assume as many electrons as protons. With the prominent exception of neutrinos, all other hadrons and leptons are present in negligible amount because they decayed long ago. Big Bang Nucleosynthesis (BBN) makes predictions that are extremely well confirmed by observations: 75% Hydrogen (single proton with only traces of deuterium) and $24.5 \pm 0.004\%$ of Helium [6] (2 protons, 2 neutrons, with traces of ^{3}He). All other elements have negligible densities as we will see in the next lecture. There are three main ways to measure Ω_b:

CMB The oscillating patterns in the *CMB power spectrum* is very sensitive to all cosmological parameters. $\Omega_b h^2$ changes the height of the acoustic peaks because it displaces the averages of the oscillations and changes the diffusion damping scale. Planck gives the impressively tight constraint (see Table 3 of [1])

$$\Omega_b h^2 = 0.02225 \pm 0.00016 \rightarrow \Omega_b \simeq 0.05 \,. \tag{3.3}$$

[12] The bound come mostly from the instrument FIRAS on board of the COBE satellite [28], which reported in 1996, $\mu < 9 \times 10^{-5}$. Recently, the ground based experiment TRIS [31] has provided a mild improvement ($\mu < 6 \times 10^{-5}$) by decreasing the degeneracy with other parameters.

Quasar absorption and the intergalactic medium Quasars are extremely luminous active galactic nuclei powered by accretion onto supermassive black holes, and they can be observed out to redshifts $z \gtrsim 7$. Their bright, nearly featureless ultraviolet continua make them excellent backlights for studying the diffuse matter distributed along the line of sight. As quasar photons propagate through the intergalactic medium (IGM), they are absorbed by neutral hydrogen at the Lyman-α resonance, corresponding to a rest-frame wavelength of

$$\lambda_\alpha = 1215.67 \text{ Å} \quad \text{or} \quad \nu_\alpha \simeq 2.47 \times 10^{15} \text{ Hz}.$$

Each intervening hydrogen cloud imprints a narrow absorption line at an observed wavelength $\lambda_{\rm obs} = \lambda_\alpha(1+z)$, forming the characteristic *Lyman-α forest* seen at higher frequencies than the quasar's intrinsic emission line. The distribution and depth of these lines trace the density and ionization state of the cosmic baryons.
Hydrodynamical simulations calibrated to the observed mean flux decrement and the one-dimensional power spectrum of the transmitted flux allow one to infer the physical baryon density parameter $\Omega_b h^2$. Analyses of large datasets such as SDSS and HIRES yield values around

$$\Omega_b h^2 \simeq 0.021 \pm 0.002,$$

in excellent agreement with the determinations from Big Bang nucleosynthesis and the cosmic microwave background (*e.g.* Planck). These measurements confirm that most of the ordinary matter at redshifts $z \sim 2$–4 resides not in galaxies or stars, but in the diffuse, photoionized IGM traced by the Lyman-α forest.

Big bang nucleosynthesis The abundance of light elements produced during *Big Bang Nucleosynthesis* (BBN) depends very sensitively on the baryon density, leading to the constraint $\Omega_b h^2 = 0.022 \pm 0.006$ [15].
As we will see, observations of the total matter density in the Universe point to a much larger fraction than a few percent, so there must be some other type of non-baryonic matter[13].

Big Bang nucleosynthesis (BBN) provides one of the earliest and most direct probes of the baryon content of the Universe. It describes the network of nuclear reactions that took place within the first few minutes after the Big Bang, when the temperature was $T \sim 10^9$ K and the Universe was dense and hot enough for nuclear fusion to occur but still simple enough to be computed from first principles. As the Universe expanded and cooled down, free neutrons and protons combined to form the lightest nuclei—deuterium (^{2}H), helium (^{3}He and ^{4}He), and trace amounts of lithium (^{7}Li). Heavier elements could not form because the density and temperature dropped too quickly for efficient reactions beyond helium, so all heavier nuclei were synthesized much later in stars and supernovae.
The predicted primordial abundances of these light elements depend sensitively on a single cosmological parameter, the baryon-to-photon ratio $\eta \equiv n_b/n_\gamma$. Since the photon number is well known from the CMB, this gives the physical baryon density $\Omega_b h^2$. In particular, the deuterium abundance is an excellent "baryometer", since its predicted value scales ap-

[13] Remember that in the cosmology slang, "matter" means a component scaling approximately as $\rho_b \propto \rho_{DM} \propto a^{-3}$.

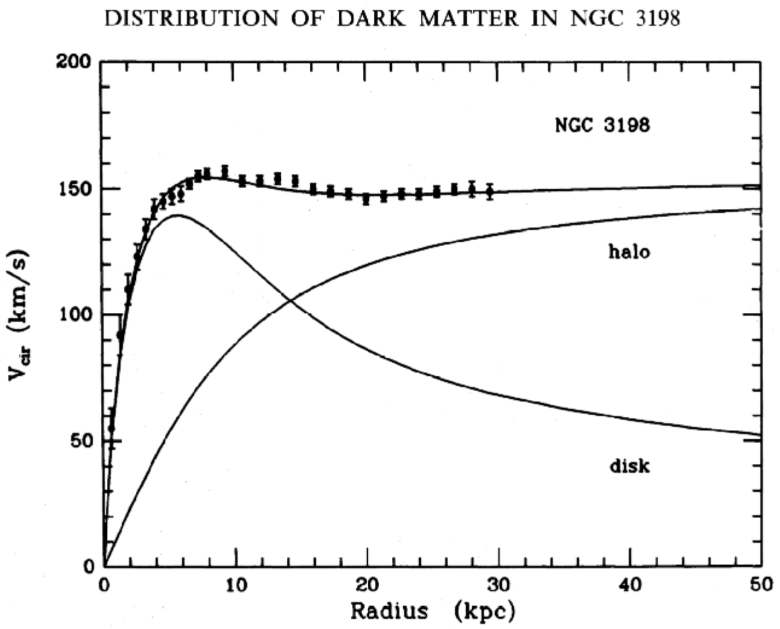


Figure 6: The plot shows the orbital velocity of stars in the galaxy NGC 3198 as function of radius. The "disk" line shows what would be expected if all the matter in the galaxy were the baryons in the (flat) galactic disk. By adding a diffuse dark matter component, a.k.a. "halo", one can reconcile predictions with observations.

proximately as (D/H) $\propto \eta^{-1.6}$. Observations of primordial deuterium in high-redshift, low-metallicity quasar absorption systems, combined with precise nuclear reaction rates, yield

$$\Omega_b h^2 \simeq 0.021 \pm 0.001,$$

in striking agreement with the value inferred from the cosmic microwave background and the Lyman-α forest. The consistency among these independent probes confirms the standard cosmological model and demonstrates that nearly all of the baryonic matter produced in the early Universe remains accounted for, though most of it resides today in diffuse gas rather than in stars.

3.3 Dark matter

The evidence for a non-baryonic component or *dark matter* comes exclusively from gravitational physics, unlike that for baryons, which we observed using the light they emit. Evidence comes from all possible scales: galaxy rotation curves, cluster counts, mass-to-light ratio and the cosmic microwave background. We will only very briefly cover the main idea of each of these fascinating topics.

Galaxy Rotation Curves In the absence of dark matter, the circular velocity $v(r)$ of a star orbiting in a galaxy of enclosed mass $M(<r)$ follows directly from Newtonian dynamics,

$$v^2(r) = \frac{G\,M(<r)}{r}.$$

If most of the mass were concentrated in the luminous disk and bulge, then for r beyond the visible edge $M(<r) \simeq$ const and hence

$$v(r) \propto r^{-1/2},$$

as in the familiar Keplerian decline observed for planetary orbits in the Solar System.
However, spectroscopic observations of spiral galaxies show that the rotation curves remain approximately flat, $v(r) \simeq v_0 \simeq$ const, out to radii several times larger than the optical disc. From the relation above this implies

$$M(<r) \propto r,$$

meaning that the total gravitating mass continues to rise linearly with radius even where little luminous matter is seen. For example, a typical spiral galaxy with an outer rotation speed of $v_0 \simeq 200$ km s^{-1} at $r \simeq 30$ kpc requires an enclosed mass

$$M(<r) \simeq \frac{v_0^2 r}{G} \simeq 3 \times 10^{11}\,M_\odot,$$

while the stellar and gas content contribute only about $5 \times 10^{10}\,M_\odot$. The excess mass therefore exceeds the visible component by a factor of several, indicating that galaxies are embedded in extended halos of non-luminous *dark matter*. This conclusion, first established through the pioneering measurements of Rubin and Ford in the 1970s, remains one of the most direct dynamical evidence for dark matter in the Universe.

Galaxy Clusters Galaxy clusters provide another classical dynamical probe of dark matter. In a relaxed cluster, the motions of member galaxies are governed by the gravitational potential of the entire cluster, allowing an estimate of the total mass through the virial theorem,

$$2\langle T\rangle + \langle U\rangle = 0, \qquad \Rightarrow \qquad M_{\rm tot} \simeq \frac{3\,\sigma_v^2\,R}{G},$$

where σ_v is the one-dimensional velocity dispersion of the galaxies and R is the characteristic cluster radius. In the 1930s, Fritz Zwicky applied this method to the Coma cluster, measuring $\sigma_v \simeq 1000$ km s^{-1} and $R \simeq 1$ Mpc, which implied a total mass $M_{\rm tot} \sim 10^{15}\,M_\odot$. However, the visible galaxies and hot intracluster gas account for less than about ten per cent of this mass. Zwicky coined the term *dunkle Materie* (dark matter) to describe the missing mass required to bind the cluster gravitationally. Modern X-ray and gravitational-lensing observations confirm his conclusion: galaxy clusters are dominated by dark matter, which shapes their potential wells and provides an essential test of cosmological models on the largest bound scales.

Mass-to-Light Ratios An alternative way to infer the presence of dark matter is through measurements of the *mass-to-light ratio*, defined as

$$\Upsilon \equiv \frac{M_{\rm tot}}{L},$$

where $M_{\rm tot}$ is the total gravitating mass and L the total luminosity of a system, both expressed in solar units. For stellar populations like the Sun or typical galaxies, the expected ratio from visible matter alone is $\Upsilon_\star \sim 1$–$5\ M_\odot/L_\odot$, depending on the stellar population and wavelength band. However, dynamical mass estimates obtained from galaxy rotation curves, velocity dispersions, and gravitational lensing show much larger values: spiral galaxies have $\Upsilon \sim 10$–50, galaxy clusters reach $\Upsilon \sim 200$–400, and the Universe as a whole requires an average $\Upsilon_{\rm cosm} \sim 3000$ to match the critical density. The steadily increasing mass-to-light ratio from stellar to cluster scales demonstrates that the majority of the gravitating mass is non-luminous, consistent with the existence of a dominant dark matter component pervading galaxies and clusters alike.

Cluster Counts The abundance of galaxy clusters as a function of redshift provides a sensitive probe of cosmological parameters. The number of clusters observed above a given mass threshold in a redshift interval dz and solid angle $d\Omega$ is proportional to the comoving volume element,

$$\frac{dV}{dz\, d\Omega} = \frac{c}{H(z)}\, d_A^2(z),$$

which depends on the Hubble parameter $H(z)$ and the angular diameter distance $d_A(z)$. The predicted number density also depends on how many clusters actually form, which in turn depends on how fast or slow cosmic structures grow from the initial seeds. This is encoded in the matter power spectrum amplitude σ_8 and the matter density parameter Ω_m. Consequently, cluster counts primarily constrain a combination of these two parameters, often expressed as

$$\sigma_8\, \Omega_m^\alpha \simeq \text{const}, \qquad \alpha \approx 0.3 - 0.5,$$

where the precise degeneracy exponent depends on the details of the mass function and redshift range considered.

Cosmic Microwave background The CMB was emitted when baryons where ionized and therefore evolved very differently from dark matter. The CMB temperature-temperature power spectrum is sensitive to Ω_{DM}, e.g. through the relative height of odd and ever peaks and the size of the diffusion damping. The current bound from Planck is $\Omega_{DM} h^2 = 0.1198 \pm 0.0015$ leading to $\Omega_{DM} \simeq 0.0267$ [1].
The matter power spectrum possesses small oscillations because the baryons were oscillating with the photons before decoupling, unlike dark matter. These *Baryon Acoustic Oscillations* pin down the ratio Ω_b/Ω_m giving a consistent measurement

3.4 Neutrinos

Neutrinos (see [24, 33, 37, 38] for a review) are the lightest fermions in the standard model and come in three families: ν_e, ν_μ and ν_τ. They carry no electric charge or "color" and interact

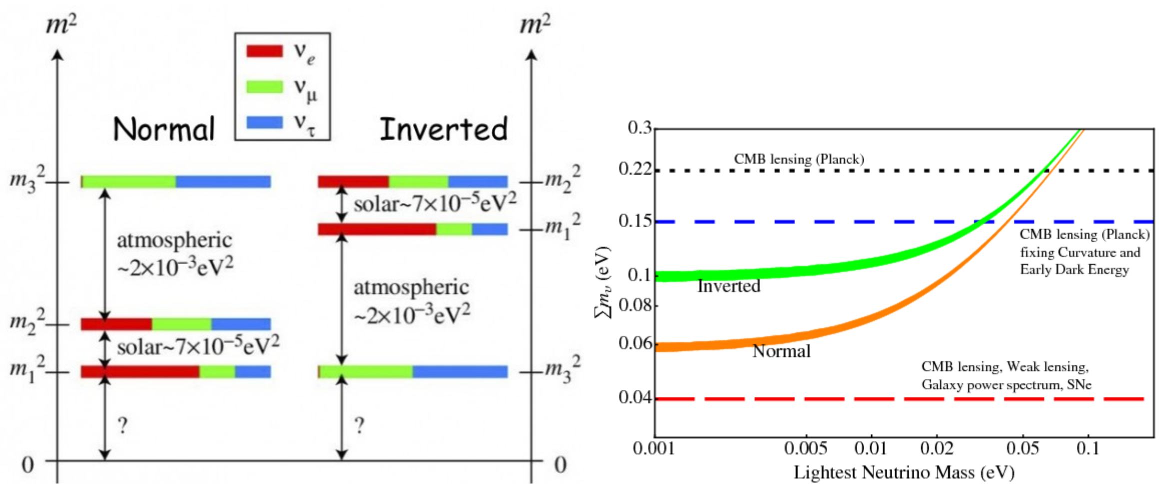


Figure 7: Left: the normal and inverted scenarios for the neutrino hierarchy. Right: the total neutrino mass as function of the yet unknown mass of the lightest neutrino. The current bounds are shown in the black dotted and blue dashed lines, while the red long-dashed line represents the expected future sensitivity. This in particular shows that the sum of neutrino masses will be detected in the near future.

weakly being part of an SU(2) doublet together with each family of charged left-handed leptons, namely the electron, muon or tau. These neutrinos have well-defined weak charge but they are not energy eigenstates. The linear relation between charge and energy eigenstates is

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}, \qquad (3.4)$$

where $s_{ij} \equiv \sin\theta_{ij}$ and $c_{ij} \equiv \cos\theta_{ij}$. Hence, the free propagation of neutrinos is determined by three masses, three mixing angles θ_{ij} and one CP-violating phase δ. Neutrino oscillations imply that at least two neutrinos have non-zero mass (Nobel prize 2015) [42]:

$$\Delta m^2_{21} \;=\; (7.9^{+1.0}_{-0.8}) \times 10^{-5}\ \mathrm{eV}^2\,, \qquad |\Delta m^2_{31}| = (2.2^{+1.1}_{-0.8}) \times 10^{-3}\ \mathrm{eV}^2\,. \qquad (3.5)$$

What these measurements cannot determine is the overall scale of neutrino masses as the sign of Δm^2_{31}. The latter uncertainty implies that there are two possible mass ordering for the three eigenstates, as shown in the left panel of Fig. 7. At present, the tightest bounds on the *sum of neutrino masses* come from cosmology. Combining CMB anisotropies with Baryon Acoustic Oscillations (BAO) gives (see 6.4 of [1])

$$\sum_{i=1}^{3} m_i < 0.17\,\mathrm{eV}\,. \qquad (3.6)$$

There is no explanation for the neutrino mass in the standard model and various models have been proposed. It is also still not known whether neutrinos are their own antiparticle, namely they are Majorana fermions, or not, namely they are Dirac fermions like the electron-positron.

With the large improvement in the precision of cosmological observations, we have now many different probes that will be able to detect neutrino masses and determine the correct hierarchy in the next 5 to 10 years! See the righthand panel of Fig. 7 for a summary of current and future bounds.
Unlike their mass, the abundance of cosmological neutrino, which is sometimes called CNB or CνB for Cosmological Neutrino Background, has been observed via CMB anisotropies. The actual constraint is often quoted in terms of the effective number of neutrinos N_{eff}. Standard model predicts $N_{eff} = 3.04$ [43], which is fully compatible with the current CMB constraints $N_{eff} = 3.04 \pm 0.18$ (see 6.4 of [1]).
If neutrinos are massive, when they are fully non-relativistic their energy density is simply given by $\rho_\nu = m_\nu n_\nu$, with n_ν their conserved number density. One then finds (see P.3.3)

$$n_\nu = \frac{3}{11} n_\gamma = \frac{6\zeta(3)}{11\pi^2} T_\gamma^3 \simeq 113\,\mathrm{cm}^{-3}\,, \tag{3.7}$$

$$\Omega_\nu = \frac{\rho_\nu}{\rho_{cr}} = \frac{\sum_{i=1}^3 m_i}{94\,h^2\,\mathrm{eV}} \quad (m_\nu \neq 0)\,. \tag{3.8}$$

Neutrinos were originally proposed to explain the entirety of dark matter but they are too light and one finds $\Omega_\nu \leq 0.4\%$. Nevertheless, neutrinos do cluster[14] and they do produce small effects on structure formation. A large number of experiments aims at detecting these effects in the next decade.

3.5 Dark energy

In the late 90's evidence began to accumulate that $\ddot{a}(t_0) > 0$, i.e. the current expansion of the Universe is accelerating. The discovery was announced by two groups: High-Z Supernova Search Team [52] and the Supernova Cosmology Project [49], both of which got the Nobel prize in 2011. Supernovae of Type 1a are exploding stars whose progenitor is a small and compact start called a white dwarf in a binary system (i.e. orbiting another, usually larger star). SN1a are standard candles so their intrinsic luminosity should be approximately the same. In practice, to achieve the high accurace needed for cosmology these need to be corrected for dust absorption and some "unknown" environmental dependence. We can calibrate nearby SN1A and hence know the intrinsic luminosity L. So, if we measure a SN1A, we can deduce its luminosity distance d_L, since we know L and measure the flux l in (2.24). In addition, the redshift of each supernova can be measured from emission and absorption lines in its spectrum. The resulting luminosity distance $d_L(z) = \chi(z)(1+z)$ or apparent magnitude $m-M$ as function of redshift are shown in Fig. **??**. This is somewhat analogous to the classic Hubble diagram in Fig. 1, with the remarkable difference that is extends to much further objects. For example we see supernovae at $z \sim 1$, which corresponds to a few Gpc. this is much further away than the few Mpc in Hubble's diagram. One sees that SN appear fainter in our Universe than they should appear in a matter dominated Universe. Introducing a cosmological constant, the measurements agree with predictions. Also, this estimate of the accelerated expansion agrees

[14] Clustering means to get denser or sparser around over or underdensities. Very relativistic particles, such as for example photons, do not cluster because they cannot be captured by the gravitational potential of even the largest clusters of galaxies.

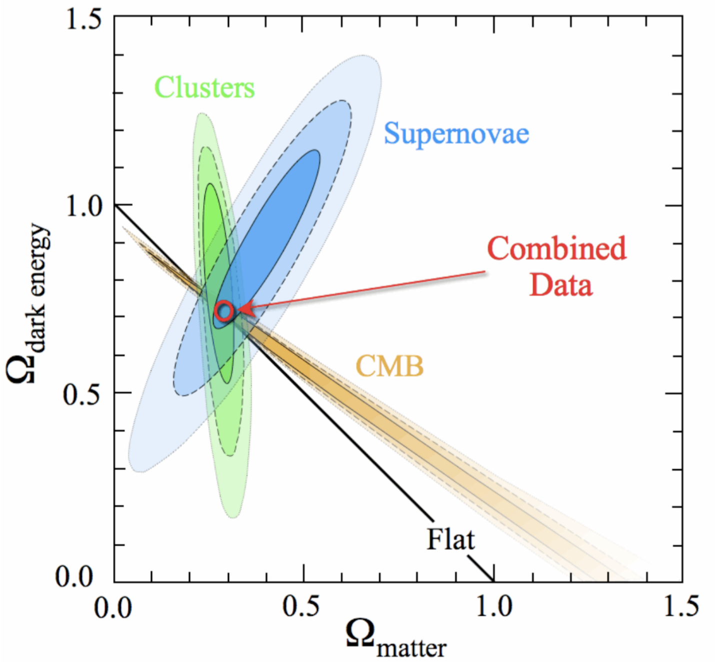


Figure 8: The concordance or standard model of cosmology. Cluster counts, supernovae and the CMB agree on a flat ($K = 0$ accelerated Universe, dominated by dark energy)

beautifully with CMB and cluster counts. This is known as the concordance model or LCDM for Λ Cold Dark Matter, and it is depicted in Fig. 8).
From the acceleration equation (1.52) we know that $\ddot{a} > 0$ implies $\rho + 3P < 0$ or equivalently $w < -1/3$. Neither matter nor radiation can produce this effect since they both obey the Strong Energy Condition, i.e. for ever future pointing time-like vector X^μ

$$\left(T_{\mu\nu} - \frac{1}{2} T^\lambda{}_\lambda g_{\mu\nu}\right) X^\mu X^\nu \geq 0 \quad \Rightarrow \quad \rho + 3p \geq 0 \quad \text{(SEC)}\,. \tag{3.9}$$

We are then forced to either change the framework within which we interpret the data, e.g. change the laws of gravity or the FLRW metric, or introduce a new constituent of the Universe: *dark energy*. A detailed study of the data shows that for dark energy to produce the accelerated expansion of the Universe, namely $\ddot{a} > 0$, we need

$$p_{DE} = -\rho_{DE}\,(1 \pm 0.05) \quad \text{and} \quad \Omega_{DE,0} \simeq 0.7 \tag{3.10}$$

Dark energy is the politically correct and over-encompassing name for all the proposed theories of late cosmic acceleration. The *cosmological constant*, *quintessence* and *modified gravity* are among the most investigated scenarios for dark energy.

Let us briefly discuss the most conservative solution to late cosmic acceleration: a cosmological constant. Diffeomorphism invariance of the EE allows for an additional constant term

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = M_{\rm Pl}^{-2} \left(T_{\mu\nu} - \Lambda_{cc} g_{\mu\nu}\right)\,, \tag{3.11}$$

which had originally been introduced by Einstein to find a static Universe, see P.3.4. Interpreting $-\Lambda_{cc} g_{\mu\nu}$ as the energy-momentum tensor of the cosmological constant, we deduce $p_{cc} = -\rho_{cc}$ and therefore[15] $w_{cc} = -1$. Notice that as the Universe expands or contracts the energy density remains constant. Because of this, the cosmological constant is also called *vacuum energy* meaning that $\rho_{cc} = \Lambda$ is an energy density associated with "empty" spacetime itself. Equivalently, the most general action compatible with the diffeomorphism covariance of GR is

$$S = \int d^4x\sqrt{-g}\left[\Lambda_{cc} + \frac{M_{\rm Pl}^2}{2}R + \mathcal{O}(R^2)\right] , \tag{3.12}$$

where additional terms such as R^2 or R^3 have more spacetime derivatives. In our conventions the spacetime constant Λ_{cc} has dimension of an energy density $[\Lambda_{cc}] = E^4$. What can we say about the value of Λ_{cc}?

The cosmological constant problem Consider General Relativity (GR) as and Effective Field Theory[16] (EFT). Since $[R] = E^2$, the theory is not renormalisable in the traditional sense, i.e. at every order in perturbation theory new operators need to be introduced with increasing number of fields and derivatives to cancel new Ultra-Violet divergences. However, just like for every EFT, at energies E well below the naive cutoff $\Lambda_{\rm cutoff} \sim M_{\rm Pl}$, there is just a finite number of counterterms needed to compute finite, renormalized observables for experiment with some finite precision ϵ. Naively, for predictions at some scale $E \ll \Lambda_{\rm cutoff}$ we need only operators of dimension Δ where $(E/\Lambda_{\rm cutoff})^{\Delta-4} \geq \epsilon$. So the theory is predictive as long as we can make independent measurements to impose all renormalization conditions on all operators of dimension Δ or less. We can then safely quantise gravity perturbatively, around some fixed classical background such as FLRW. When we couple GR to a given model of particle physics, the additional dynamics might introduce strong coupling at lower energies than $M_{\rm Pl}$. Since we have successfully tested the standard model of particle physics at accelerators, we conservatively assume TeV$< \Lambda_{\rm cutoff} < M_{\rm Pl}$. Here comes the key point. In a *natural* EFT's, every coupling constant is expected to be given by appropriate powers of the cutoff of the theory. For example Λ is expected to be the cutoff of the theory to the fourth power, and so

$$\Lambda_{cc} \simeq \Lambda_{\rm cutoff}^4 > {\rm TeV}^4 \quad \text{(natural EFT expectation!?)}\,, \tag{3.13}$$

But the late time cosmic acceleration is an Infra-Red (IR) effect as compared with typical particle physics scales. Late acceleration is associated with an energy density in the Universe of order

$$3H_0^2 M_{\rm Pl}^2 \sim (10^{-3}\times {\rm eV})^4 \ll {\rm TeV}^4 < \Lambda_{\rm cutoff}^4\,. \tag{3.14}$$

So the expectation based on naturalness is at least wrong by a factor of $(10^{15})^4$. The above considerations about the cosmological constant are usually summarised in terms of two conceptually distinct problems:

- The cosmological constant *naturalness problem* or why don't we observe a large contribution to the Universe energy budget of order $\Lambda_{\rm cutoff}^4$ >TeV^4?

[15] CFU: *What is the critical w for which we change from accelerated to decelerated expansion? Look back at the acceleration equation Eq. (1.52)*

[16] For an introductory discussion of EFT's see e.g. [51]

- The cosmological constant *fine-tuning problem* or how does the tiny dimensionless number $\Lambda^4_{\text{cutoff}}/M^2_{\text{Pl}}H^2_0 > 10^{60}$ emerge from the laws of nature?

Many models have been constructed to address these issues over the years, but there is no clear favourite so far. An ambitious observational program is underway to test many of these theories. For more details see e.g. [16, 59].

Problems for lesson 3

P.3.1 Compute the galaxy rotation curve, namely the velocity v as function of radius R, assuming that there is only baryonic matter (stars and interstellar gas, but no dark matter). You can use the Newtonian approximation and assume a Gaussian baryonic distribution $\rho(R) = \rho_0 e^{-(R/R_s)^2}$ where R_s is typically of order a few kpc, e.g. $R_s \sim 4\,\text{kpc}$ for our Milky way. Notice that the distribution of luminous matter can be deduced from the luminosity of the galaxy as function of radius. The Gaussian profile above reproduces only qualitatively the exponential decay at large radius. Qualitatively compare your result with some actual data (e.g. google image "galaxy rotation curves").

P.3.2 (Dodelson's Exercises 17 ch.2) Express the entropy density s as function of temperature T for massless bosons and fermions, assuming equilibrium and zero chemical potential. Neglecting chemical potential, show that a particle of mass m in equilibrium at $T \ll m$ gives an exponentially small contribution to the entropy $s \propto e^{-m/T}$.

P.3.3 (Dodelson's Exercises 18 ch.2) Show that the number density of one generation of neutrinos and anti-neutrinos in the Universe today is approximately

$$n_\nu = \frac{3}{11} n_\gamma \sim 100\,\text{cm}^{-3}\,. \tag{3.15}$$

P.3.4 Einstein originally introduced a cosmological constant in order to maintain a static Universe. Find out how he was proposing to realize this. Consider a Universe with matter, radiation, a cosmological constant defined by

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + g_{\mu\nu}\Lambda = \frac{1}{M^2_{\text{Pl}}} T_{\mu\nu}\,, \tag{3.16}$$

and curvature K, so that

$$H^2 + \frac{K}{a^2} = \frac{1}{3M^2_{\text{Pl}}}\left(\rho_M + \rho_R + \rho_\Lambda\right)\,, \tag{3.17}$$

with $\rho_\Lambda \equiv M^2_{\text{Pl}}\Lambda$. Find the special value $\bar{a}$ of a and of Λ such that the Universe is static. Is this static solution stable under perturbations away from $\bar{a}$?

P.3.5 Consider two supernovae, one with apparent magnitude $m = 24.3$ at $z = 0.83$ and one with $m = 16.08$ at $z = 0.026$. Neglecting error bars and assuming a flat Universe with just matter and a cosmological constant, determine the preferred value of Ω_Λ.

4 Inflation: Motivations

In this section we discuss several problems with any cosmological model in which the universe is dominated by radiation in the far past, all the way until the big bang. We will refer to this class of models collectively as the "hot big bang" model, where "hot" refers to the temperature of radiation. In particular, the root of all problems will be that most[17] of the expansion ($\dot{a} > 0$) of the universe e.g. in ΛCDM is decelerated $\ddot{a} < 0$. Decelerated expansion starts from the big bang (which in ΛCDM would happen during radiation domination) at $z \to \infty$ or $a \to 0$ and lasts all the way until dark energy takes over "recently" around $z \simeq 0.5$. First, we discuss old "background" problems, namely the horizon and curvature problems, which can be stated already for the unperturbed FLRW universe that we have studied so far. These problems were originally formulated in the 80's and have not changed much since. Second, we discuss new "perturbation" problems, namely scale invariance and phase-coherence problems, which have to do with the large amount of new data we have collected in the past 30 years, especially from the Cosmic Microwave Background (CMB). Finally, in preparation for the next lecture, we review the basic properties of the maximally symmetric spacetime with positive cosmological constant, i.e. de Sitter spacetime.

4.1 Old background problems

In the following we discuss two of the problems that were well known more than 40 years ago and pushed many cosmologists to modify the early expansion history of our universe.

Curvature problem The first background problem is that we do not observe any spatial curvature in our universe, despite the fact that curvature dilutes more slowly than radiation and matter (and in fact than anything obeying the strong energy condition) and should grow with time relatively to them. Let us see this in formulae.
Current bounds tell us that [1]

$$\Omega_K \equiv \left(\frac{K}{a^2H^2}\right) , \quad \Omega_{K,0} = 0.000 \pm 0.005 \,. \tag{4.1}$$

On the other hand, as we saw in Sec. 1, the most general homogeneous and isotropic space can have spatial curvature, i.e. $K \neq 0$. From Eq. (4.1) we see that Ω_K grows with time in an decelerated ($\ddot{a} < 0$) expanding ($\dot{a} > 0$) universe

$$\dot{\Omega}_K = -\ddot{a}\frac{2K}{\dot{a}^3} \propto -\ddot{a} \propto (\rho + 3p) \propto (1 + 3w) \,, \tag{4.2}$$

where in the second step we used the acceleration equation (1.52) to show that in an expanding universe ($\dot{a} > 0$) the Strong Energy Condition (SEC), see (3.9), implies deceleration. Since at early times in ΛCDM the universe is dominated by radiation, $w = 1/3$, we conclude that Ω_K must have been even smaller in the past[18]. In other words, extrapolating closer and closer to the big bang singularity at $a \to 0$ and $\rho \to \infty$, we are forced to assume that the initial

[17] This is measured on a log scale, i.e. the duration of a cosmological phase is measured in terms of $\log(a_f/a_i)$, where $a_{i,f}$ are the initial and final value of the scale factor.

[18] CFU: *Estimate Ω_K at the time of big bang nucleosynthesis.*

curvature was tiny, $\Omega_K(a_i) \to 0$, or equivalently the initial total density of the universe was extremely close to the critical one, $\sum_i \rho_i \to \rho_c$ (defined in 1.49). There are only three logical possibilities:

1. The curvature of the universe is zero to begin with, and so it did not grow with time. While this is a possibility in an exactly homogeneous universe, it is very unlikely to be realized in our universe because we observe non-vanishing perturbations on all scales. In particular, we measure deviations from exact FLRW of order $\Delta(\lambda) \sim 10^{-5}$ at wavelength λ of order the (physical) Hubble radius $\lambda \sim 1/H$. These perturbations are approximately scale invariant for shorter scales, $\lambda < 1/H$ and so it is natural to expect that there exists non-vanishing perturbation of a similar amplitude on superHubble scales $\lambda \gtrsim 1/H$. Such perturbations would induce a local spatial curvature of the order

$$K = \frac{\Delta(\lambda)}{\lambda^2} \quad \Rightarrow \quad \Omega_{K,0} = \frac{\Delta(\lambda)}{\lambda^2 H_0^2} \lesssim 10^{-5} \,. \tag{4.3}$$

   This argument strongly disfavours this possibility.

2. The initial conditions of the universe, as it emerged from some non-perturbative theory of quantum gravity[19], were extremely fined tuned close to Ω_K. In this scenario, the existence of the universe as we know it is a very rare fluctuation, since any larger initial value of $\Omega_K(t_i)$ would have grown to dominate the energy density of the universe and prevented the formation of galaxies and therefore life as we know it. Also not a great option, in the opinion of many.

3. The early expansion history of our universe is modified to stop Ω_K from growing as we move back in time. From (4.2) we see that this requires either $\ddot{a}, \dot{a} < 0$, i.e. an early phase of decelerated contraction, or $\ddot{a}, \dot{a} > 0$, i.e. an early phase of accelerated expansion. Since we know the current universe is expanding (recall Hubble's law), the first of these options requires to *bounce* i.e. to transition from $\dot{a} \propto H < 0$ to $\dot{a} \propto H > 0$. Achieving the bounce in a controlled construct is still an open problem and the many proposed models have a series of pathologies. Therefore we focus on an early phase of accelerated expansion, a.k.a. cosmological *inflation*, in the rest of these notes.

Summarising, to avoid fine tuned initial conditions for the universe, we postulate the existence of a primordial phase of accelerated expansion, $\ddot{a}, \dot{a} > 0$, called inflation.

Horizon problem A second background problem of the hot big bang model is that the homogeneity of the observed universe on large scales is at odds with the decelerated expansion history. In fact, cosmological observations of far away objects allow us to see regions in the

[19]CFU: *Strictly within GR, K is just a parameter, not a dynamical variable, and so there in no physical perturbation that can make $\Omega_K = 0$ unstable. On the other hand, GR is most likely just a low-energy effective description of some UV-complete theory of quantum gravity, and it is at least plausible that $\Omega_K = 0$ might be unstable within that larger, yet unknown theory. Perhaps a more concrete example is bubble nulceation. Instanton solutions are known in which a new universe nucleates from a single point [14]. To respect the isometries of the system the new universe must have some negative curvature. It is not known whether bubble nucleation and the ensuing ideas about the multiverse play a role in the history of our own universe, and the discussion among experts continues.*

past that are much larger than the particle horizon at the time. Any mechanism attempting to explain the observed homogeneity in a causal way then necessarily violates causality, leading the horizon problem.

To see this quantitatively, recall that the comoving distance (see Sec. 2) between two generic times t_1 and t_2 with $a_1 = a(t_1) < a(t_2) = a_2$ is found to be

$$\chi(a_1, a_2) \equiv \int_{a_1}^{a_2} \frac{da}{a^2 H} = \frac{1}{a_1 H_1} \frac{2}{3w+1} \left[\left(\frac{a_2}{a_1} \right)^{(3w+1)/2} - 1 \right] , \tag{4.4}$$

where we assumed $w \neq -1/3$. Then, the distance of an object at redshift $1 + z = a^{-1}$ from us at $a = a_0 = 1$ is given by

$$\chi(a, 1) \equiv \int_{a}^{a_0} \frac{d \log a}{aH} = \frac{1}{H_0} \frac{2}{3w+1} \left[1 - a^{(3w+1)/2} \right] , \tag{4.5}$$

Imagine now to look out in the night sky in opposite directions and detect a pair of antipodal object, each sending us radiation with the same[20] redshift z. The relative comoving distance $\Delta\chi$ between the objects is just $2\chi(a, 1)$. To simplify the algebra, let us neglect dark energy[21] and so $w > -1/3$ (In ΛCDM $w \in \{0, 1/3\}$) and assume $a \ll 1$. Then

$$\Delta\chi(a, 1) \simeq 2 \times \frac{1}{H_0} \frac{2}{3w+1} \simeq \frac{\mathcal{O}(1)}{H_0} , \tag{4.6}$$

Recall that the redshift of these objects is $1 + z = 1/a$, and so we conclude that high redshift objects $z \gg 1$ are at a distance of order the Hubble radius today H_0^{-1}, almost independently of z.[22] Since this is a comoving distance between objects at fixed comoving position (i.e. far away object are in the Hubble flow), it does not depend on time. Let us compare now this distance with the comoving particle horizon in a hot big bang model, i.e. extrapolating radiation domination all the way to $a_i = 0$. Recall that the comoving particle horizon[23] $x_{\text{p.h.}}$ is the comoving distance traveled by light since the beginning of time τ_i, namely $x_{\text{p.h.}}(a) \equiv \chi(a_i, a)$. Notice that $x_{\text{p.h.}}$ depends on the integral in (4.4) over the whole history of the universe, as opposed for example to the Hubble radius r_H, which carries information about a single instant of time. Recall also that for $w > -1/3$, or equivalently decelerated expansion $\ddot{a} < 0$ (as it is the case for radiant and dust), one can safely take $a_i \to 0$ and so $x_{\text{p.h.}}(a)$ equals the comoving

[20] This assumption is clearly not necessary, but it allows us to avoid obfuscating ideas with indices.

[21] CFU: *Check that this does not affect the argument at all.*

[22] CFU: *Using the Hubble law, show that the Hubble radius H^{-1} represents the physical distance beyond which a comoving object moves away from us faster than the speed of light, namely $\partial_t x_{phy} > c = 1$.*

[23] This is simply related to the physical particle horizon $d_{\text{p.h.}}$ of (2.21) by $a x_{\text{p.h.}}(a) \equiv d_{\text{p.h.}}(a)$

Hubble radius[24] times an order one number[25]

$$x_{\text{p.h.}}(a) = \frac{1}{aH}\frac{2}{3w+1} \simeq \frac{1}{aH}\mathcal{O}(1) \simeq r_H(a)\mathcal{O}(1) \quad \text{(decelerated)}\,. \tag{4.7}$$

Assuming decelerated expansion since the big bang, one finds

$$\frac{\Delta\chi(a)}{x_{\text{p.h.}}(a)} \simeq 2\,\frac{aH}{a_0H_0} \simeq 2\left(\frac{1}{a}\right)^{(3w+1)/2} \gg 1 \quad \text{(decelerated)}\,. \tag{4.8}$$

We just learnt that, in an ever decelerating universe, by observing far away objects ($1/a = 1+z \gg 1$) we are actually probing scales much larger than the particle horizon at that time. In practice, one can reach $a = (1+z)^{-1} \sim 0.1$ with quasar and $a \sim z^{-1} \sim 10^{-3}$ with Cosmic Microwave Background (CMB) photons. In both cases, the observed physical properties (e.g. density of quasars, temperature and polarization of the CMB) are the same in the opposite directions in average. We conclude that, in the absence of accelerated expansion in our past, the mechanism responsible for this observed statistical isotropy must violate causality. This is the *particle horizon problem*.
Conversely, for a phase of accelerated expansion, $\ddot{a} > 0$ or $w < -1/3$ (such as during dark energy or inflation) during a period $a \in \{a_i, a_f\}$, the result is divergent as $a_i \to 0$:

$$x_{\text{p.h.}}(a_f) = \frac{1}{a_fH_f}\frac{2}{|3w+1|}\left[\left(\frac{a_f}{a_i}\right)^{|3w+1|/2} - 1\right] \tag{4.9}$$

$$\simeq \frac{1}{a_fH_f}\frac{2}{|3w+1|}\left(\frac{a_f}{a_i}\right)^{|3w+1|/2} \gg r_H \quad \text{(accelerated)}\,. \tag{4.10}$$

In the extreme case $w \simeq -1$ (inflation), H is approximately constant and $x_{\text{p.h.}}$ asymptotes to the constant value

$$x_{\text{p.h.}} \to \frac{1}{a_iH_i} \quad \text{(inflation)}\,. \tag{4.11}$$

Yet again, if we want to keep causality as a guiding principle, we must postulate a phase of accelerated expansion $\ddot{a}, \dot{a} > 0$ in the early universe[26], or a phase of decelerated contraction $\ddot{a}, \dot{a} < 0$, which entails severe technical problems.

[24] Recall the comoving *Hubble radius* r_H, which is defined as

$$r_H \equiv \frac{1}{aH} = \frac{1}{\dot{a}} = \frac{a^{(3w+1)/2}}{H_0} \qquad \text{(single fluid)}\,,$$

where in the third equality we used the solution of the Friedmann equation for a single fluid with $p = w\rho$ and constant w. As usual, the physical Hubble radius is simply $r_{H,\text{phys}} = ar_H = H^{-1}$. In the literature, r_H is often referred to as *Hubble "horizon"*. This is a misnomer since neither $(aH)^{-1}$ nor its physical cousin H^{-1} represent a horizon in the usual sense of GR. This nomenclature is widely spread and not harmful as long as one is aware of the subtleties. In these notes, we will try hard to use the expressions"Hubble radius" or just "Hubble scale" instead of "Hubble horizon".

[25] CFU: *Show that if two different decelerated phase follow each other (radiant and matter domination in our universe), the contribution from the latter dwarfs that of the former.*

[26] CFU: *Prove that, during decelerated expansion, $\ddot{a} < 0$, perturbations "re-enter" the Hubble horizon, in the sense that*

$$\frac{\partial}{\partial t}\left(\frac{\lambda_{phy}}{H^{-1}}\right) < 0\,,$$

and viceversa.

Figure 9: The plot shows the evolution of the comoving distance and the particle horizon in the phase of early accelerated and late decelerated expansion. The large growth of the particle horizon during inflation ensures that it is causally possible for any point in the current observable universe today to exchange information with any other.

The horizon problem is well summarised by the plot in Fig. **??**, which shows the time evolution of x_H and $x_{\rm p.h.}$ for our universe. The ordinate represents time and is parameterized by the number of *e-foldings* of expansion

$$dN \equiv H dt = d \log a \quad \Rightarrow \quad N = \log a + \text{const} \,. \tag{4.12}$$

We have chosen the integration constant so that $N = 0$ separates early accelerated expansion, i.e. inflation, from late deceleration, i.e. radiation and matter domination. The abscissa of the upper and lower panels indicates physical and comoving scales respectively. The black lines represent Hubble radius, while ... Diagonal, thin, red lines represent the physical wavelength λ_{phy} or comoving wavelength λ of some monochromatic perturbation.

4.2 New perturbation problems*

There are problems with the hot big bang models that were not known 40 years ago because the data was not good enough. We believe these "new" problems must play an important role in guiding us towards a theory of the early universe. Since these new problems deal with perturbations they are best understood after we have learned to describe the inhomogeneous universe in Part II. Here we attempt to give a cursory and intuitive introduction to these problems. This subsection is not examinable.

Phase coherence problem As we saw discussing the horizon problem, by observing distant objects at $z \gg 1$, we can see scales much larger than the Hubble radius at the time. Our universe does have perturbations already on these superHubble scales, i.e. with wavelength $\lambda > 1/H$. What's really remarkable is that all these superHubble perturbations we have observed appear to oscillate in exact synchronicity: they have all the same phase! This is the *phase coherence* of cosmological perturbations, which give rise to the distribution of galaxies in the universe today. In an ever decelerating universe, the Hubble radius and the particle horizon are the same up to an unimportant order one factor. In this case then phase coherence is observed even on scales much larger than the particle horizon. This is a problem because on these super-horizon scales no causal mechanism can be devised to synchronize the phases and so their coherence becomes a very unlikely coincidence. This strongly suggests that there was a primordial phase, before the hot big bang, during which perturbations were produced and synchronized, rather than being generated at late times, during the hot big bang. Let us see this more in detail.
In the CMB, for each direction $\hat{n}$ of the sky ($\hat{n}\cdot\hat{n} = 1$), we observe both temperature fluctuations $\Delta T(\hat{n}) \equiv T(\hat{n}) - \bar{T}$ around the average temperature $\bar{T}$, and a specific type of photon polarization called E-mode and denoted by $E(\hat{n})$. Because of the isotropy of the universe on large scales, it is convenient to decompose fields on the sphere into spherical harmonics

$$X(\hat{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm}^{X} Y_{lm}(\hat{n}) \quad \Rightarrow \quad a_{lm}^{X} \quad = \quad \int d^2\hat{n} \, X(\hat{n}) \, Y_{lm}^{*}(\hat{n}) \,, \tag{4.13}$$

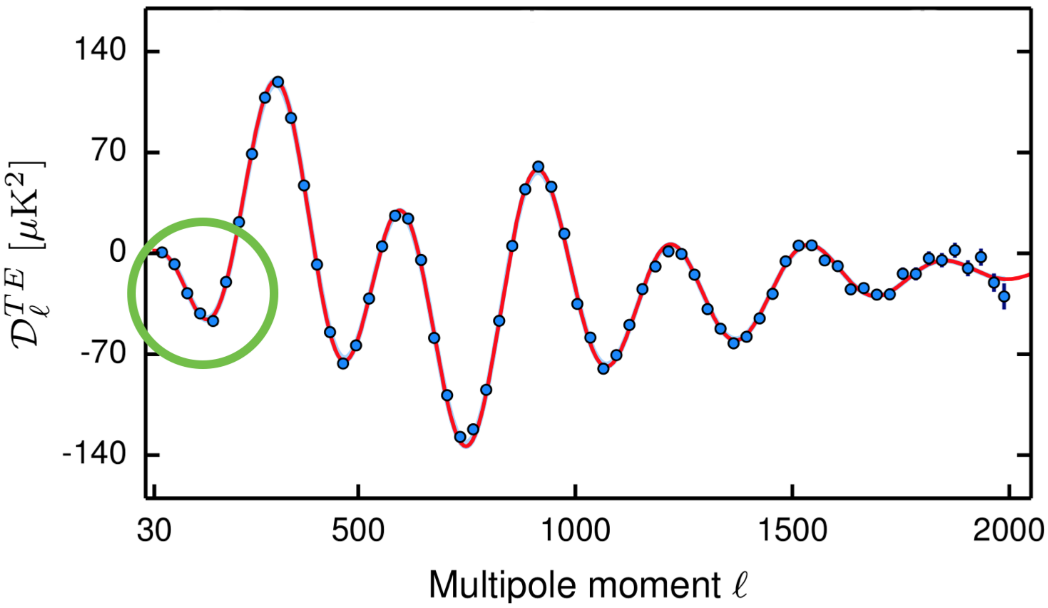


Figure 10: The plot shows the cross-correlation between CMB temperature T and E-mode polarization [1]. The *anti*-correlation around $l \sim 100$ shows that superHubble perturbations at the time of last scattering exist and they oscillate with coherent phases.

where $X = \{\Delta T, E\}$. The isotropy of the universe tells us that different values of m correspond to independent realisations of the universe. Using the ergodic theorem, we can then approximate quantum or stochastic averages, which we can compute from the theory side with angular averages, which can be observed experimentally

$$\langle \hat{\mathcal{O}}_1 \hat{\mathcal{O}}_2 \dots \rangle \quad \sim \quad \frac{1}{(2l+1)} \sum_m a_{lm}^{\mathcal{O}_1} a_{lm}^{\mathcal{O}_2} \dots . \tag{4.14}$$

$$\text{theory} \quad \leftrightarrow \quad \text{observations} \tag{4.15}$$

For example, the correlation between ΔT and E can be obtained observationally from the observed spherical harmonic coefficients

$$\langle a_{lm}^T a_{lm}^E \rangle = \frac{1}{(2l+1)} \sum_m a_{lm}^T a_{lm}^E \equiv C_l^{TE} . \tag{4.16}$$

It is customary to plot the quantity $\mathcal{D}_l^{ET} \equiv l(l+1)C_l^{ET}$ to make the figure more visible. This correlation was measured most recently by the Planck satellite is shown in Fig. 10 as function of the multipole l. The green circle draws your attention to the negative cross-correralation for $l \lesssim 100$.

Let us see how we can interpret this feature on the theory side. At cartoonish level, temperature fluctuations are a measurement of dimensionless density fluctuations of the photon-electron-baryon plasma, while the polarization is a measurement of the divergence of the plasma velocity

$v(\mathbf{x}, t)$ at the spacetime point of origin $(\mathbf{x}, t)$ of the CMB photon[27].

$$\frac{\Delta T(\mathbf{x}, t)}{\bar{T}} \sim \delta \equiv \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \,, \qquad E(\mathbf{x}, t) \sim \partial_i v^i(\mathbf{x}, t) \,. \tag{4.17}$$

One therefore finds

$$\langle a_{lm}^T a_{lm}^E \rangle \sim \langle \delta \, \partial_i v^i \rangle \,, \tag{4.18}$$

We now need to specify the stochastic properties of δ and $\partial_i v^i$, so that we can compute this average. Consider the simplest possible toy model: a single, monocromatic (sound) wave

$$\delta(\mathbf{x}, t) = A \cos(\mathbf{k} \cdot \mathbf{x}) \cos(\omega t + \phi) \,, \tag{4.19}$$

where ω is some fixed frequency, A is the amplitude and ϕ the phase. To mimic the real calculation we should be doing in a quantum mechanical universe, we will assume that A and ϕ are some random variables drawm from some distribution to be specified. Using the linearised continuity equation

$$\dot{\delta} + \partial_i \left[(1 + \delta) \, v^i \right] \simeq \dot{\delta} + \partial_i v^i = 0 \quad \text{(fluid continuity eq.)} \,, \tag{4.20}$$

we can compute the velocity as well

$$\partial_i v^i(\mathbf{k}, t) = -\dot{\delta}(\mathbf{x}, t) = \omega A \cos(\mathbf{k} \cdot \mathbf{x}) \sin(\omega t + \phi) \,. \tag{4.21}$$

Now we need to assume something about the probability distribution that governs A and ϕ. For this, let us consider the comoving particle horizon at the time the CMB was emitted, the "last scattering" of photon, at redshift $z_{LS} \simeq 1100$. We know from (4.7) that in a decelerating universe this is approximately the same as the comoving Hubble radius $(aH)_{LS} \simeq 4 \times 10^{-3}$ Mpc^{-1}, corresponding to CMB multipoles of approximately $l_{LS} \sim \tau_0 k_{LS} \simeq 70$. Therefore observations on $l \lesssim l_{LS}$ effectively measure perturbations that were super-horizon at the time of emission. In addition, perturbations with $l_{LS} < l < 150$ have spent less than one Hubble time H^{-1} inside the Hubble radius. Since their typical frequency is also of the order of H, they have evolved little from their initial value on superHubble scales. It seems then reasonable to assume that the distribution of ϕ is not peaked around any specific value, since no causal process could have chosen one over another. We will then tentatively assume a flat distribution for $\phi \in \{0, 2\pi\}$, i.e. incoherent, uncorrelated phases. Then the cross-correlation vanishes,

$$\langle \delta \partial_i v^i \rangle \;\propto\; \langle AA \rangle \langle \cos(\omega t + \phi) \sin(\omega t + \phi) \rangle \tag{4.22}$$

$$= \langle AA \rangle \int_0^{2\pi} d\phi \cos(\omega t + \phi) \sin(\omega t + \phi) = 0 \,, \tag{4.23}$$

where non-random variables such as ω and $\cos(\mathbf{k} \cdot \mathbf{x})$ can be factored outside the average. Since this correlator was our proxy for C_l^{TE}, which is instead observed to be negative and far from zero in Fig. 10, we conclude that the initial superHubble phases were not random but rather *coherent*. In other words, any two perturbations (with the some fixed wavenumbers $|\mathbf{k}| = |\mathbf{k}'|$

[27] This is of course an extreme oversimplification. The different physical effects are discussed a bit more in detail in Box **??**.

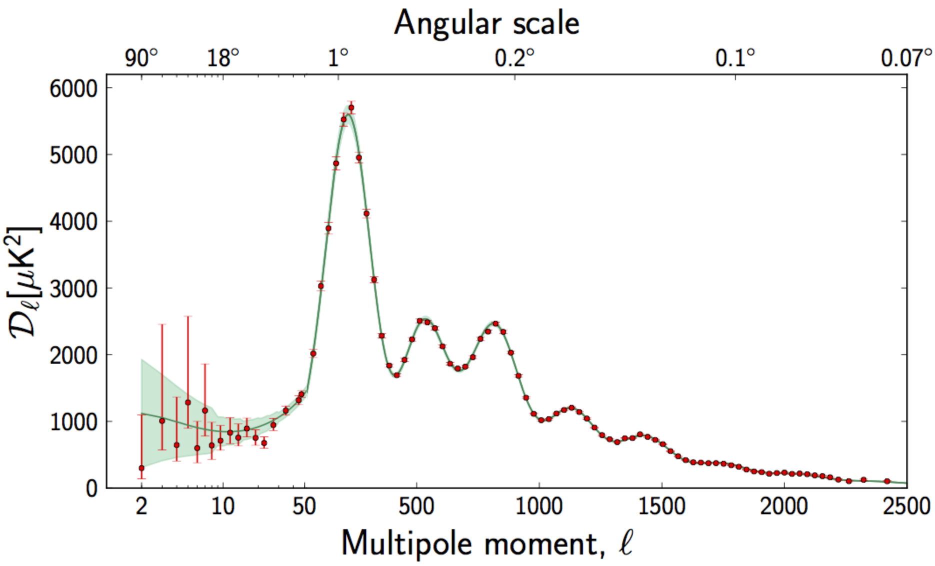


Figure 11: The angular power spectrum of CMB temperature anisotropies as measured by the Planck satellite.

corresponding to the same l) must have been synchronised at some early time before the hot big bang.

One last piece of evidence as to how the synchronisation might have taken place is the negative sign of the correlation. Gravitational collapse is often quoted to make "the rich richer and the poor poorer". This alludes to the fact that, when pressure is negligible, the leading (growing) mode of linearized gravitational collapse consists of a flow away from underdense regions into overdense ones. In formulae

$$\delta > 0 \quad \Rightarrow \quad \dot{\delta} > 0 \quad \Rightarrow \quad \partial_i v^i \sim -\dot{\delta} < 0 \,, \tag{4.24}$$

and viceversa, where in the last step we used the (non-relativistic, linear) continuity equation[28]. Notice that, even if one started with some different initial conditions, say with completely uncorrelated δ and $\partial_i v^i$, always in the absence of pressure, this mode will eventually dominate. Therefore, we would not be surprised to find anti-correlations on scales that have spend some sizable amount of time evolving inside the Hubble radius in the absence of pressure. On the other hand, the negative ET correlation on large scales, $l < 150$, tells us that the coherent superHubble perturbations where already in the "growing" mode, even though there was not enough time for any late-time dynamics to select this mode. Some sort of gravitational collapse must have started already in the very early universe.

Scale invariance problem The second and last problem with the perturbed universe is the surprising fact that the amplitude of perturbations observed in our universe is approximately the same (within 4%) on all cosmological scales (about 3 orders of magnitude $10^{-4} - 10^{-1}$

[28] CFU: *Check that the addition of the linear relativistic correction (e.g. in Newtonian gauge) does not alter the sign of* $\partial_i v^i$.

Mpc−1). This remarkable feature of what we can now call *primordial perturbations* goes under the name of (approximate) *scale invariance*[29] . The mathematical statement is that for every $\lambda \in \mathbb{R}$ and $n \in \mathbb{N}^+$, a field ϕ obeys scale invariance iff[30]

$$\langle \phi(\mathbf{x}_1)\phi(\mathbf{x}_2)\dots\phi(\mathbf{x}_n)\rangle = \langle \phi(\lambda\mathbf{x}_1)\phi(\lambda\mathbf{x}_2)\dots\phi(\lambda\mathbf{x}_n)\rangle \,, \tag{4.26}$$

where all the fields are evaluated at the same time[31]. Scale invariance is most evident in the large scales ($l \lesssim 40$) of the CMB temperature angular power spectrum, i.e. the average (or quantum correlator)

$$C_l^{TT} \equiv \frac{1}{2l+1}\sum_l a_{lm}^T (a_{lm}^T)^* = \langle a_{lm}^T (a_{lm}^T)^*\rangle \,. \tag{4.27}$$

From Fig. 11, we see that on large scales or small multipoles $l \ll 70$, where we can neglect the acoustic oscillations of the photon-electron-baryon plasma, the angular power spectrum C_l is well approximated by $\mathcal{D}_l = l\,(l+1)\,C_l = \text{const}$.
By using and abusing the flat sky approximation[32] , one finds

$$\langle \delta T(\hat{n})\delta T(\hat{n}')\rangle \simeq \int dl^2 dl'^2\, e^{i(\mathbf{l}\cdot\mathbf{n}+\mathbf{l}'\cdot\mathbf{n}')}\,\langle \delta T(\mathbf{l})\delta T(\mathbf{l}')\rangle \tag{4.29}$$

$$\simeq \int dl^2 dl'^2\, e^{i(\mathbf{l}\cdot\mathbf{n}+\mathbf{l}'\cdot\mathbf{n}')}\,\langle a(\mathbf{l})a(\mathbf{l}')\rangle \tag{4.30}$$

$$\simeq \int dl^2\, e^{i\mathbf{l}\cdot(\mathbf{n}-\mathbf{n}')}\, C_l \,. \tag{4.31}$$

Since $C_l \sim l^{-2}$, one recognises in the last line the solution of Poisson's equation[33] with a uniform constant source. By appropriately regulating the divergence, the solution is a constant, i.e. independent of $\mathbf{n}-\mathbf{n}'$, so the primordial correlation function of $\mathcal{R}$ is independent of scale (distance $|\mathbf{n}-\mathbf{n}'|$) as advertised. An analogous derivation goes through using the large scales of the matter power spectrum (see right panel of Fig. P.5.4), but we leave this to the ambitious reader.

[29] CFU: *Primordial perturbations are most easily discussed in terms of the curvature perturbation $\mathcal{R}$, which are time independent on superHubble scale. In this sense, the initial conditions can be though of as correlators in a $(0+3)$-dimensional field theory. In this Euclidean interpretation correlators are fully conformal invariant*

[30] CFU: *Derive the equivalent statement for the correlators of the Fourier transform of the field $\phi(\mathbf{k})$. In particular, for the two-point function in Fourier space, a.k.a. the power spectrum, you should find*

$$\langle \phi(\mathbf{k})\phi(\mathbf{k}')\rangle = (2\pi)^3 \delta_D\left((\mathbf{k}+\mathbf{k}'\right)\frac{C}{k^3}\,, \tag{4.25}$$

for some constant C.

[31] Beware that this is Cosmology lingo. In other fields, such as Conformal Field Theory, sometimes the term scale invariance is used to refer to the invariance under scaling of time as well as space in the correlators.

[32] CFU: *The flat sky approximation corresponds to the substitution*

$$\frac{\delta T}{T}(\hat{n}) = \sum_{lm} a_{lm} Y_{lm}(\hat{n}) \to \Theta(\mathbf{n}) = \int dl^2\, e^{i\mathbf{l}\cdot\mathbf{n}}\,\Theta(\mathbf{l})\,, \tag{4.28}$$

where the coordinates of the sphere $\hat{n} = \{\theta, \phi\}$ are approximated by euclidean 2d coordinates $\mathbf{n} = \{n_1, n_2\}$. This is valid as long as we consider only a small portion of the sphere (sky).

[33] CFU: *The mathematically inclined reader can proceed to perform the integral directly by using polar coordinates and the residue theorem. It is useful to include a small tilt $C_l \propto l^{-2+\epsilon}$ to regulate the result.*

Box 4.1 Invariance under translations and rotations Consider the most general homogeneous and isotropic spaces, namely an FLRW space. If all other relevant background quantities are also homogeneous and isotropic, then all primordial correlators must be left invariant by the generators of spatial translations and rotations. In real space, these are

$$P_i : -\partial_i \quad \text{and} \quad R_{ij} : -\left(x_i \partial_j - x_j \partial_i\right) , \tag{4.33}$$

and act on the argument of each perturbation ϕ (assumed to be a scalar for simplicity) as in

$$\sum_{a=1}^{n} \frac{\partial}{\partial \mathbf{x}_a} \langle \phi(\mathbf{x}_1)\phi(\mathbf{x}_2)\ldots\phi(\mathbf{x}_n)\rangle \stackrel{!}{=} 0 , \tag{4.34}$$

$$\sum_{a=1}^{n} \left(x_a^i \frac{\partial}{\partial x_a^j} - x_a^j \frac{\partial}{\partial x_a^i} \right) \langle \phi(\mathbf{x}_1)\phi(\mathbf{x}_2)\ldots\phi(\mathbf{x}_n)\rangle \stackrel{!}{=} 0 . \tag{4.35}$$

The general solution of the first constraint is that the correlator only depends on $n-1$ variables, for example $\mathbf{x}_a - \mathbf{x}_1$ for $a = 2, \ldots n$. The generators acting on Fourier space correlators are

$$P_i : -k_i \quad \text{and} \quad R_{ij} : -\left(k_i \partial_j - k_j \partial_i\right) , \tag{4.36}$$

and therefore

$$\sum_{a=1}^{n} \mathbf{k}_a \langle \phi(\mathbf{k}_1)\phi(\mathbf{k}_2)\ldots\phi(\mathbf{k}_n)\rangle \stackrel{!}{=} 0 , \tag{4.37}$$

$$\sum_{a=1}^{n} \left(k_a^i \frac{\partial}{\partial k_a^j} - k_a^j \frac{\partial}{\partial k_a^i} \right) \langle \phi(\mathbf{k}_1)\phi(\mathbf{k}_2)\ldots\phi(\mathbf{k}_n)\rangle \stackrel{!}{=} 0 . \tag{4.38}$$

The first condition is satisfied if the correlator is proportional to $\delta(\sum_a^n \mathbf{k})$, while the second requires it to depend only on the rotationally invariant contractions $\mathbf{k}_a \cdot \mathbf{k}_b$.

One would like to see scale invariance emerging from some (scaling) symmetry of the primordial physics that generated perturbations. A very simple and elegant solution is found by assuming that, during some primordial era, the background spacetime was well approximated by *de Sitter space* (dS) in flat slicing (see Sec. 4.3)

$$ds^2 = \frac{-d\tau^2 + dx^i dx^j \delta_{ij}}{\tau^2 H^2} , \tag{4.32}$$

for some constant Hubble parameter H. One of the ten isometries of this maximally symmetric spacetime is the *dilation* symmetry[34]

$$\tau \to \lambda \tau , \quad \mathbf{x} \to \lambda \mathbf{x} . \tag{4.39}$$

If all other non-gravitation background quantities depend very weakly on time, then Eq. (4.39) is an approximate symmetry of the full theory and primordial correlators must be invariant under it. Following [17], it is then immediate to see scale invariance arise. In Fourier space, under the transformation Eq. (4.39), a field scales as $\phi(\mathbf{k}, \tau) \to \phi(\mathbf{k}/\lambda, \lambda\tau)$ so the power spectrum must take the form in Eq. (4.25) up to an arbitrary function $F(k\tau)$, which must be zero

[34] CFU: *Prove this assertion using the definition Eq. (1.58)*

if the field under consideration is constant, as it is the case for $\mathcal{R}$ on superHubble scales.

It is useful to prove this simple result using a more cumbersome but also more powerful formalism. It is easiest work again in Fourier space and introduce the following notation

$$\langle\phi(\mathbf{k}_1)\phi(\mathbf{k}_2)\dots\phi(\mathbf{k}_n)\rangle = (2\pi)^3\,\delta_D\left(\sum_{b=1}^{n}\mathbf{k}_b\right)\langle\phi(\mathbf{k}_1)\phi(\mathbf{k}_2)\dots\phi(\mathbf{k}_n)\rangle'\,. \tag{4.40}$$

Then the generator of dilations in real space[35] is

$$D:\ -\tau\partial_\tau - x^i\partial_i \quad (\text{real space})\,, \tag{4.42}$$

acting on *each* field in the correlator. When acting on primed Fourier-space correlators $\langle\dots\rangle'$, the generator becomes[36]

$$D:\ -3+\sum_{a=1}^{n}(3-\tau_a\partial_{\tau_a})+k_a\frac{\partial}{\partial k_a} \quad (\text{Fourier space})\,. \tag{4.43}$$

The desired scale invariance is obtained by requiring that D leaves correlators of $\mathcal{R}$ invariant. Since *these*$\mathcal{R}$ is conserved on superHubble scales, we can drop the time derivatives and find

$$\left[3(n-1)+\sum_{a=1}^{n}k_a\frac{\partial}{\partial k_a}\right]\langle\mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\dots\mathcal{R}(\mathbf{k}_n)\rangle' \stackrel{!}{=} 0\,. \tag{4.44}$$

For the power spectrum[37] $P_\mathcal{R}(k)\equiv\langle\mathcal{R}(\mathbf{k})\mathcal{R}(\mathbf{k}')\rangle'$, this gives

$$\left[3+k\frac{\partial}{\partial k}+k'\frac{\partial}{\partial k'}\right]P_\mathcal{R}(k)\stackrel{!}{=}0 \quad\Rightarrow\quad P_\mathcal{R}(k)=\frac{C}{k^3}\,, \tag{4.45}$$

for some constant C. Summarizing, the observed scale invariance of the primordial power spectrum follows directly from the dilation isometry of de Sitter space.

4.3 de Sitter spacetime

De Sitter spacetime (dS) is one of three maximally symmetric spacetimes, together with Anti-de Sitter (AdS) and Minkowski space. Recall from Sec. 1, that maximally symmetric spaces in $D=d+1$ spacetime dimensions have $D(D+1)/2$ isometries[38]. Therefore, in our $(3+1)$-dimensional world, dS has 10 Killing vectors. It arises as a solution of Einstein equations in the presence of a cosmological constant

$$R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R+g_{\mu\nu}\Lambda=0\,. \tag{4.46}$$

[35] CFU: *Check that indeed $\xi^\mu=\{-\tau,-x^i\}$ is a Killing vector for the dS metric in Eq. (4.39), namely it solves*

$$\mathcal{L}_\xi g_{\mu\nu}=-\left(\nabla_\mu\xi_\mu-\nabla_\mu\xi_\mu\right)=0\,. \tag{4.41}$$

where $\mathcal{L}$ is the Lie derivative. Convince yourself that this equation is equivalent to Eq. (1.58).

[36] CFU: *If you desire reproducing this, keep in mind that the -3 in front comes from the Dirac delta we factored out in Eq. (4.40), the $+3$ comes from the Fourier transform in each coordinate and we used the identity $\mathbf{k}\cdot\partial_\mathbf{k}=k\partial_k$.*

[37] CFU: *Using Eq. (4.44) derive the scaling of any n-point function.*

[38] CFU: *This is easily remembered as the dimension of the $(d+1)$-dimensional Poincaré group $\mathbb{R}^{(d,1)}\times SO(d,1)$ or as that of the $(D+1)$-dimensional Lorentz group $SO(D,1)$.*

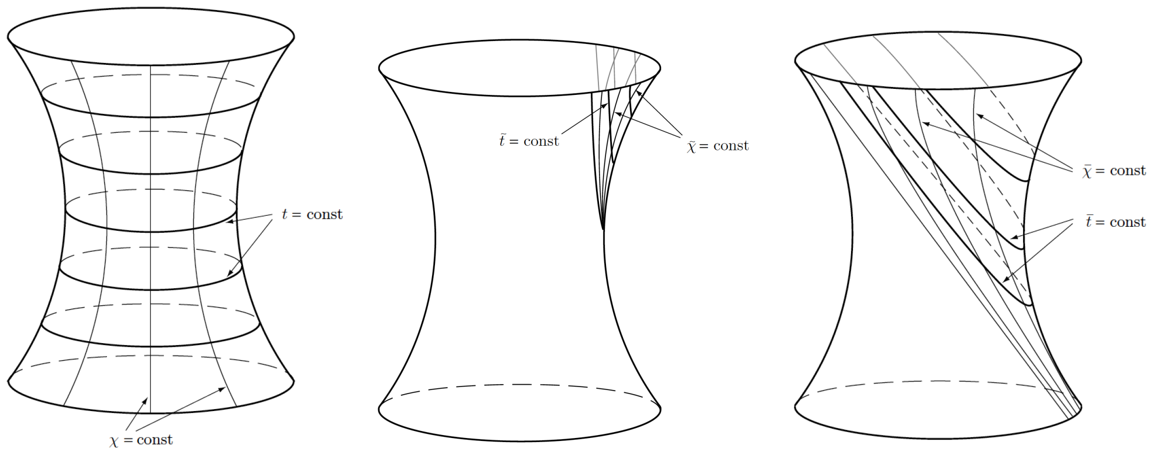


Figure 12: The three time slicing of dS space. From left to right they are closed, open, and flat slicing. Figure adapted from [46].

The trace of this expression (for $D > 2$) tells us $R = \Lambda 2D/(D-2)$ and therefore dS is an Einstein manifold, namely the Ricci tensor is proportional to the metric[39]

$$R_{\mu\nu} = \frac{2\Lambda}{D-2} g_{\mu\nu} \,. \tag{4.48}$$

dS in D-dimensions can be defined as a codimension one, hyperbolic surface in $(D+1)$-dimensional Minkowski space, defined by[40]

$$-\left(X^0\right)^2 + \sum_{a=1}^{D} X^a X^a = L^2 \,, \quad \text{with} \quad \Lambda = \frac{(D-2)(D-1)}{2L^2} \,, \tag{4.49}$$

where L is the dS radius. The dS hyperboloid is invariant under $(D+1)$-dimensional Lorentz transformations, but not translations, namely the group $SO(D,1)$. While the $(D+1)$ Minkowski coordinates of Eq. (4.49) are useful because they transform linearly under this $SO(D,1)$ isometry group, they are clearly redundant. There are three common ways to define D non-redundant coordinates (see [61] for other useful coordinates), which differ in how dS is sliced into constant time hypersurfaces. All three slicings can be thought of as intersecting the dS hyperboloid in Eq. (4.49) with a one-parameter family of D-dimensional hyperplanes:

- If the vector perpendicular to the planes is time-like with respect to the $(D+1)$ metric, namely the planes are more "horizontal" than 45 degrees, their intersection with the hyperboloid has a finite volume. Without loss of generality, one can choose the planes to be horizontal. The intersection leads to the circles on the left-hand side of Fig. 12.

[39] Actually, in $d = 3$ the full Riemann tensor is also given in terms of the metric

$$R_{\mu\nu\rho\sigma} = \frac{R}{12}\left(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}\right) \,. \tag{4.47}$$

[40] CFU: *Notice that the sign of* $+L^2$ *is such that the hyperboloid lies outside the light cone.*

This is called the *closed slicing* of dS because the constant time hypersurfaces of dS are hyper-spheres, with positive spatial curvature and finite volume.

- Analogously, a family of "vertical" planes provides the *open slicing*, with constant-time hypersurfaces of negative spatial curvature and infinite volume.
- The case in between, namely 45 degrees planes, has flat constant-time hypersurfaces of infinite volume. This slicing is commonly used for inflation, which dilutes curvature and makes it negligibly small. The flat-slicing metric in normal and conformal time is[41]

$$ds^2 = -dt^2 + e^{2Ht}dx^2 = \frac{-d\tau^2 + dx^2}{\tau^2 H^2}\,, \tag{4.50}$$

related to the Minkowski coordinates by (here $i = 1, \ldots, d-1$)

$$X^0 = L\sinh(Lt) - \frac{1}{2}\frac{x^i x_i}{L}e^{-Lt}\,, \quad X^i = x^i e^{-Lt}\,, \quad X^d = L\cosh(Lt) - \frac{1}{2}\frac{x^i x_i}{L}e^{-Lt}\,. \tag{4.51}$$

Finally, it is useful to consider combinations of dS coordinates that are invariant under dS isometries. The simplest one requires two points and can be thought of as an *invariant distance*. Using the (redundant) $(D+1)$ Minkowski coordinates, this distance is obviously

$$|X - X'|^2 = \left(X - X'\right)^\mu \eta_{\mu\nu} \left(X - X'\right)^\nu\,, \quad (\mu = 0, 1, \ldots, d)\,. \tag{4.52}$$

Since the two points X and X' lie on the dS hyperboloid, $|X|^2 = |X'|^2 = L^2$, so the only part of this distance that actually depends on their position is $X^\mu \eta_{\mu\nu} X'^\nu$. It is therefore convenient to define the invariant distance as

$$D(X; X') \equiv -X^0 X'^0 + X^i X'^i \quad (i = 1, \ldots, d)\,, \tag{4.53}$$

$$D(t, x^i; t', x'^i) \equiv \cosh\left(Ht - Ht'\right) - \frac{|x - x'|^2}{2H^2} e^{-H(t+t')}\,, \tag{4.54}$$

$$D(\tau, x^i; \tau', x'^i) \equiv \frac{\tau^2 + \tau'^2 - |x - x'|^2}{2\tau\tau'}\,, \tag{4.55}$$

for the different sets of coordinates.

[41] CFU: *Derive the relation between the Hubble parameter H and the dS radius L.*

5 Single-field slow-roll inflation

The problems encountered in the previous section suggested we need a prolonged phase of accelerated expansion, with a background close to dS, which we will call *inflation* [32]. In this section, we move beyond these kinematical considerations and discuss the dynamics of inflation.
As we saw in the previous section around Eq. (4.46), a cosmological constant Λ supports a dS solution. However, as the name suggests, the cosmological *constant* does not change with time and the dS phase would be eternal, and could not be connected to the universe as we know it. There is an easy fix: let us introduce a clock ϕ that "turns off" Λ after some time so that the dS phase can indeed stop when desired. we will call this clock-dependent cosmological non-constant $V(\phi)$, to avoid confusing it with the cosmological constant Λ. We can now proceed in two different directions:

1. We can simply specify some function $\phi(t)$ and obtain the desired inflationary background. Naively, this breaks explicitly the diffeomorphism invariance upon which GR is built and seems to introduce a time-dependent function by hand. On a second thought, a gauge symmetry[42] can never be really broken (as the Stückelberg trick teaches) and the choice of time in GR is arbitrary anyways. This approach, made popular by [13], is very effective (pun intended) for model-independent discussion, to highlight the role of symmetries and finally to make connection with observations. On the other hand, it requires a higher level of abstraction than the alternative.

2. We can insist that $\phi(t)$ is the solution of some diff-invariant theory. The simplest choice, as we will see shortly, is a single, canonical scalar field minimally coupled to gravity. An advantage of this point of view is that it provides an important stepping stone to understand the origin of inflation within a UV-complete theory of gravity, such as string theory. This second approach is more intuitive and pedagogical, and so more appropriate for this introductory course.

5.1 Prolonged quasi-de Sitter expansion

The horizon, curvature and phase coherence problems taught us that we should postulate the existence of an early phase of accelerated expansion $\ddot{a}, \dot{a} > 0$, which we call *inflation*. Let us reformulate this as

$$\frac{\ddot{a}}{a} = \dot{H} + H^2 = H^2 (1 - \epsilon) > 0 \,, \tag{5.1}$$

where we have introduced the *first Hubble slow-roll parameter*

$$\epsilon \equiv -\frac{\dot{H}}{H^2} \,, \tag{5.2}$$

which is a dimensionless measure of the time variation of H. Using (1.51) it is easy to see that for a single-fluid universe $\epsilon = 3(1 + w)/2$. From (5.1), we recognise that acceleration requires $\epsilon < 1$ (or $w < -1/3$, as we knew from (1.52)). Also, as long as the matter sector

[42] CFU: *GR can indeed be thought of as a gauge symmetry, with spacetime varying Lorentz transformations.*

satisfies the Null Energy Condition (see Box 1), $\epsilon > 0$ (or $w > -1$). Observations of both of the CMB and of Large Scale Structures (LSS) probe cosmological scales over roughly three orders of magnitude[43]and observe approximate scale invariance up to percent corrections (see the scale invariance problem P.5.4). A detailed study of cosmological perturbations shows that scale invariance follows very generically if the spacetime background during inflation is close to de Sitter spacetime, i.e. H is approximately constant. Quantitatively, we will therefore be interested in $0 < \epsilon \ll 1$ (or $w \sim -1$) during inflation.
Let us estimate how long inflation has to last to address the problems discussed in the previous section. A necessary condition to solve the horizon problem is that the particle horizon is larger than the observable universe today. In terms of comoving quantities

$$x_{\text{p.h.}} > r_H = \frac{1}{a_0 H_0} \quad (\text{horizon problem})\,. \tag{5.3}$$

It is convenient to multiply both sides by the Hubble radius at the end of inflation. This is the time when the early acceleration expansion stops and the decelerated hot Big Bang starts. We will call this time *reheating* since this is when the energy is transferred from the inflationary sector to Standard Model particles. If we indicate the comoving Hubble radius by $r_{H_{\text{reh}}} = (a_{\text{reh}} H_{\text{reh}})^{-1}$ and use (4.11) for the particle horizon during a quasi de Sitter expansion, we find

$$\frac{a_{\text{reh}} H_{\text{reh}}}{a_i H_i} > \frac{a_{\text{reh}} H_{\text{reh}}}{a_0 H_0}\,, \tag{5.4}$$

where a_i indicates the beginning of inflation. There is great uncertainty about the time of reheating. We are going to parameterize this uncertainty using the temperature of the plasma of Standard Model particle at that time

$$3 M_{\text{Pl}}^2 H_{\text{reh}}^2 = g_* \frac{\pi^2}{30} T_{\text{reh}}^4\,, \tag{5.5}$$

where $g_* \sim 100$, but the precise value will not matter given the much large uncertainty in T_{reh}. Also, since the temperature of photon has approximately evolve at $T \sim 1/a$ until now[44], we can estimated $a_{\text{reh}} \sim T_{CMB,0}/T_{\text{reh}}$. Then the right-hand side of (5.4) is

$$\frac{a_{\text{reh}} H_{\text{reh}}}{a_0 H_0} \simeq 4 \times 10^{21} \left(\frac{T_{\text{reh}}}{10^{10}\text{GeV}} \right)\,. \tag{5.6}$$

The actual reheating temperature may dramatically differ from the reference temperature 10^{10}GeV, and a reasonable range of uncertainty is $T_{\text{reh}} \in \{1 - 10^{15}\}$ GeV. It is convenient

[43] CFU: *On the large-scale end, both CMB and LSS probe subHubble scales (although LSS surveys up to date have still a rather small volume and so give weaker constraints than CMB on the largest scales). On the short-scale end, CMB anisotropies are cut-off by the thickness of the last scattering surface and diffusion (a.k.a. Silk-) damping to scales of about .2× Mpc^{-1}. LSS in principle extend to shorter scales, but our lack of understanding of non-linear and baryonic physics limits our current ability to extract cosmological information from scales smaller than about* 0.2× Mpc^{-1}*. Currently, both CMB and LSS probe a similar window of scales* $\{10^{-4} - 10^{-1}\}$ Mpc^{-1}*. There is hope to enlarge this "CMB/LSS window" towards smaller scales with the CMB spectrum and 21 cm.*

[44] This neglects the changes in g_* around mass thresholds, but these again lead to small changes in the final result.

to re-express the duration of inflation on the left-hand side of (5.4) in terms of *efoldings* of expansion, defined by

$$dN \equiv \frac{da}{a} = H dt \quad \Rightarrow \quad N_2 - N_1 = \log\left(\frac{a_2}{a_1}\right) . \tag{5.7}$$

Taking the log of (5.4) we finally find

$$\Delta N_{\text{infl}} > 50 + \log\left(\frac{T_{\text{reh}}}{10^{10}\text{GeV}}\right) , \tag{5.8}$$

and so $\Delta N_{\text{infl}} \in \{25 - 60\}$. I'll often use $\Delta N_{\text{infl}} \sim 50$ for numerical estimates.

We observe approximate scale invariance for about 7 of the total ΔN_{infl} efoldings of expansion, but it is natural to assume that $\epsilon \ll 1$ remains to be valid during most of inflation. To quantify this, let us re-write the definition of ϵ and generalise it to the second and higher order Hubble slow-roll parameters[45]

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = -\partial_N \ln H \,, \tag{5.9}$$

$$\eta \equiv \frac{\dot{\epsilon}}{H\epsilon} = \partial_N \ln(\epsilon) \,, \tag{5.10}$$

$$\xi_{n\geq 3} \equiv \partial_N \ln \xi_{n-1} \,, \tag{5.11}$$

with $\xi_2 \equiv \eta$ and where we used $dN = Hdt$ from (5.7). Then, the Taylor expansion of ϵ around some reference time N_* is

$$\epsilon(N) - \epsilon(N_*) = \left.\frac{\partial \epsilon}{\partial N}\right|_{N_*} (N - N_*) + \left.\frac{\partial^2 \epsilon}{\partial N^2}\right|_{N_*} \frac{(N - N_*)^2}{2} + \mathcal{O}\left(\partial_N^3 \epsilon\right) \tag{5.12}$$

$$= \epsilon \left[\eta \,(N - N_*) + \eta \xi_3 \frac{(N - N_*)^2}{2} + \mathcal{O}\left(\eta^3, \eta^2 \xi_3, \eta \xi_3 \xi_4, \epsilon\right)\right] , \tag{5.13}$$

where all the slow-roll parameters are evaluated at N_*. The requirement that ϵ does change much during inflation is then $\eta \Delta N_{\text{infl}}, \xi_n \eta \Delta N_{\text{infl}} < 1$ and so

$$\epsilon, \eta, \xi_n \ll 1 \quad \text{(slow-roll inflation)} . \tag{5.14}$$

Note that, under the simplistic assumptions that the Taylor above expansion approximates $\epsilon(N)$ during most of inflation and that $\eta \sim \xi_n$, one can think of η^{-1} as the approximate duration of inflation in efoldings. ask how such e

5.2 Single field inflation

In the previous subsection, we have characterised the expansion history during inflation. We now want to ask how such an expansion history can emerge dynamically, from solving the equations of motion. To try to mimic a cosmological constant, we were led to consider the

[45] CFU: *Notice that all slow-roll parameters are dimensionless.*

Box 5.1 Non-canonical scalar fields A canonical scalar field has a simple quadratic kinetic term with one spacetime derivative per field, as in (5.15). We easily imagine more general but still covariant possibilities. The most generic one with at most one derivative per field is a generic function $P(X,\phi)$ of ϕ and the kinetic term $X \equiv -\partial_\mu\phi\partial^\mu\phi/2$. The homogeneous equations of motion are then

$$\ddot{\phi}\,(P_X + 2XP_{XX}) + 3H\dot{\phi}P_X + (2XP_{X\phi} - P_\phi) = 0\,, \tag{5.17}$$

while the Friedmann and acceleration equation read

$$3M_P^2H^2 = 2XP_X - P\,, \quad -M_P^2\dot{H} = XP_X\,. \tag{5.18}$$

These theories can give rise to slow-roll inflation and sometime go under the name of k-inflation [30] or simply "P-of-X " theories. An interesting subclass of these theories are those with an exact "shift symmetry" $\phi \to \phi + c$ resulting in $P = P(X)$, without any ϕ dependence. In flat space these always admit a solution $X = \text{const}$ (see Prob. P.5.3) and describe the low-energy effective theory of superfluids [60]. When minimally coupled to gravity, there are no slow-roll solutions [26] but if there is a point X_s where $\partial_X P(X)|_{X_s} = 0$, then there is an exact de Sitter solution (see Prob. P.5.3).

action of scalar field coupled to gravity. A minimally coupled[46], canonical (see Box P.5.3) scalar field is the simplest option

$$S = -\int \sqrt{-g}\frac{1}{2}\left[M_{\rm Pl}^2 R + \partial_\mu\phi\partial^\mu\phi + 2V(\phi)\right]\,, \tag{5.15}$$

where the potential $V(\phi)$ is an arbitrary function. The energy-momentum tensor (A.37) is then[47]

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left[\frac{1}{2}\partial_\mu\phi\partial^\mu\phi + V(\phi)\right]\,. \tag{5.16}$$

This takes the same form as the energy-momentum tensor of a perfect fluid (see Eq. (1.34)), under the following identifications[48]

$$\rho = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi + V(\phi)\,, \tag{5.19}$$

$$p = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\,, \tag{5.20}$$

$$u_\mu = \frac{\partial_\mu\phi}{\sqrt{-\partial_\mu\phi\partial^\mu\phi}}\,. \tag{5.21}$$

[46] Minimal coupling mean that we should write down a Lorentz invariant Lagrangian and then simply couple it to gravity with the substitutions $d^4x \to d^4x\sqrt{-g}$ and $\partial_\mu \to \nabla_\mu$. This does not capture non-minimal couplings such as for example $Rf(\phi)$ or $R^{\mu\nu\rho\sigma}\partial_\mu\phi\partial_\nu\phi\partial_\rho\phi\partial_\sigma\phi$

[47] CFU: *Compute this from the definition of $T_{\mu\nu}$*

[48] CFU: *Notice that the perfect fluid ansatz, Eq. (1.34), is more general that a single scalar field. For example, how many functions of space (initial conditions) does one need to fully specify a solution $\phi(\mathbf{x},t)$? and how many to specify $\delta(\mathbf{x},t)$ and $\mathbf{u}(\mathbf{x},t)$? Consider carefully the order of time derivatives in the equations of motion of the two systems. As we discuss in Sec. P.5.4, a scalar field maps bijectively to a perfect superfluid rather than a fluid.*

Let us focus on the homogeneous background dynamics. It is useful to specify the fluid parameterisation to the case $\phi = \phi(t)$,

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi)\,, \quad p = \frac{1}{2}\dot{\phi}^2 - V(\phi)\,, \quad u_\mu = \{1, \mathbf{0}\}\,. \tag{5.22}$$

The equation of motion for ϕ following from Eq. (5.15) are simply $\Box\phi = 0$ with the d'Alambert operator defined in Eq. (A.6). It needs to be supplemented with the Friedman equation, Eq. (**??**), to give a closed system of equations. Since we will be interested in accelerated expansion, which dilutes spatial curvature, we will set $K = 0$ in the following. For homogeneous configurations one finds[49]

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0\,, \quad \text{(background)} \tag{5.23}$$

While the first and last terms are very familiar from Newton's law, the middle term[50] represents a genuinely relativistic effect. This is sometimes called *Hubble friction* and always opposes changes in ϕ, slowing down the field. The system is closed using the Friedmann equation

$$3H^2 M_{\rm Pl}^2 = \frac{1}{2}\dot{\phi}^2 + V(\phi) \quad \text{(background)}\,. \tag{5.24}$$

For almost any potential these EOMs cannot be solved exactly[51]. On the other hand, as will see shortly, general approximate solutions are available in the regime most relevant for observations (quasi dS). Before proceeding, notice that, by taking the time derivative of Eq. (5.24) and using Eq. (5.23), one finds the very useful *exact* relation

$$-\dot{H} M_{\rm Pl}^2 = \frac{1}{2}\dot{\phi}^2\,. \tag{5.25}$$

5.3 Potential slow-roll parameters

The Hubble slow-roll parameters in (5.9)=(5.11) express in a simple and compact way the necessary requirements of an extended inflationary phase. On the other hand, their dependence on the properties of the scalar field that drives the expansion remains implicit: given some $V(\phi)$, one needs to solve the full dynamics to find $H(t)$. We will now study an approximation scheme to evaluated them more directly.

[49] CFU: *Recall that, as consequence of diffeomorphism invariance, Einstein's equations generically imply the equations of motion of matter (see e.g. Sec. 19.6 of [9]). In practice, the Bianchi identities imply the conservation of $T_{\mu\nu}$. Check these statements for a homogeneous scalar field. What happens when $\dot{\phi} = 0$? Convince yourself that gravity would erroneously think that $\phi(\mathbf{x}, t) = C$ is a solution for any C, even if $\phi = C$ is not a minimum of the potential. Ponder then on the quote from [45] (Sec. 20.6):*

> *Electromagnetism has the motto, "I count all the electric charge that's here". All that bears no charge escapes its gaze. "I weigh all that's here" is the motto of spacetime curvature. No physical entity escapes this surveillance."*

Apparently, cosmological constants do escape its surveillance.

[50] CFU: *Convince yourself that, unless the numerical coefficient is exactly 3, namely the number of space dimensions, this EOM cannot follow directly from a Lagrangian.*

[51] CFU: *The Hamilton-Jacobi formalism can be used to find the right scalar potential $V(\phi)$ that gives rise to some (restricted) class of exact solutions as discussed in Box 2.*

Box 5.2 The Hamilton-Jacobi formalism and exact solutions Following [39] and references therein, one can divide both sides of Eq. (5.25) by $\dot{\phi}$ to find

$$2H_{,\phi}M_{\rm Pl}^2 = \dot{\phi}\,, \tag{5.26}$$

where the time dependence of H has been traded for its ϕ dependence, $H(t) = H(t(\phi))$. Then the Friedmann equationEq. (5.23) can be re-written as

$$3H^2M_{\rm Pl}^2 = V + 2\left(H_{,\phi}\right)^2 M_{\rm Pl}^2\,. \tag{5.27}$$

One can then choose some function $H(\phi)$ and find the potential V form this algebraic equation. The first order differential equation Eq. (5.26) can be solved to find $\phi(t)$ and hence $H(t)$.

In the hope to find some easily calculable slow-roll parameters, one might define the *potential slow-roll parameters*

$$\epsilon_V \equiv \frac{M_{\rm Pl}^2}{2}\left(\frac{V'}{V}\right)^2\,, \quad \eta_V \equiv M_{\rm Pl}^2\frac{V''}{V}\,, \quad \xi_{3V} \equiv M_{\rm Pl}^4\frac{V'V'''}{V^2}\,, \tag{5.28}$$

and the higher orders will not be relevant for us[52]. The relation between these and the Hubble parameter can be derived by repetitively differentiating the Friedmann equation (5.24) (and using (5.25) and the definition of ϵ)

$$V = (3-\epsilon)\,H^2M_{\rm Pl}^2 \tag{5.29}$$

with respect to time and using the chain rule $\dot{V} = V'\dot{\phi}$. For example, assuming $\dot{\phi} > 0$ one finds the *exact* expressions

$$\epsilon_V = \frac{\epsilon(\eta - 2\epsilon + 6)^2}{4(\epsilon-3)^2}\,, \quad \eta_V = \frac{\eta(\eta + 2\xi_3 + 6) - 2(5\eta + 12)\epsilon + 8\epsilon^2}{4(\epsilon - 3)}\,. \tag{5.30}$$

Naively it looks like things got even more complicated. But as long as all the Hubble slow-roll parameter appearing here are small, we can find the approximate and much simpler relations

$$\epsilon \simeq \epsilon_V\,, \quad \text{and} \quad \eta \simeq 4\epsilon_V - 2\eta_V\,. \tag{5.31}$$

Notice from their definitions in Eq. (5.28), that the potential slow-roll parameter only depend on $V(\phi)$. This is in general not sufficient to know the solution of the EOM[53], Eq. (5.23), since one still has to impose two initial conditions (ϕ_i and $\dot{\phi}_i$). So what these parameters tell you is that there exist some choice of initial conditions that support and extended phase of inflation, but they do not tell you whether a given solution of the EOM does it or not. In practice, for many classes of potential the inflationary trajectory is a local attractor in phase space, so after some time, the approximation in Eq. (5.31) becomes very good. Beware though that this statement does not hold in general and in principle one needs to consider each case individually.

[52] CFU: *Higher order potential slow-roll parameters can be defined by asking that lower order ones do not change much in one efolding (or one Hubble time).*

[53] CFU: *Take from example a constant potential $V(\phi) = \bar{V}$, so that $\epsilon_V = \eta_V = 0$. The set up some initial $\dot{\phi}_i \neq 0$. Convince yourself that nevertheless ϵ and η can be very large, depending on $\dot{\phi}_i$ and V.*

5.4 Slow-roll inflation

The assumption that slow-roll parameters are small allows to find approximate solutions to the EOM. We will see that the definitions in Eq. (5.28) emerge quite naturally.
For ease of calculation and further convenience, it is useful to introduce a shorter name for the canonical kinetic term

$$X \equiv -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu \quad \overset{\text{background}}{\longrightarrow} \quad X = +\frac{1}{2}\dot{\phi}^2\,. \tag{5.32}$$

Then the relevant background equations become

$$\rho = X + V\,, \quad p = X - V\,, \quad \text{and} \quad \dot{X} + 6HX + V'\dot{\phi} = 0\,, \tag{5.33}$$

where the last equation is just the continuity equation, which is equivalent to the EOM Eq. (5.23) multiplied by $\dot{\phi}$ (see also footnote 49). Making use of Eq. (5.25), the condition $\epsilon \ll 1$ tells us that the Friedmann equation, Eq. (5.24), is dominated by the potential term V and we can neglect the kinetic term X,

$$\epsilon = -\frac{\dot{H}}{H^2} = \frac{X}{H^2} = \frac{3X}{V+X} \ll 1 \quad \Rightarrow \quad X \ll V\,, \tag{5.34}$$

and so

$$3M_{\rm Pl}^2H^2 \simeq V\,. \tag{5.35}$$

It is then straightforward to derive the *exact* relation

$$\eta = \frac{\dot{\epsilon}}{\epsilon H} = 2\epsilon + \frac{\dot{X}}{XH}\,. \tag{5.36}$$

Since $\epsilon, \eta \ll 1$ we learn that (assuming $\dot{\phi} \neq 0$)

$$\dot{X} \ll XH \quad \Rightarrow \quad 2\ddot{\phi} \ll \dot{\phi}H\,, \tag{5.37}$$

and so we can neglect the acceleration term $\ddot{\phi}$ in Eq. (5.23) (or $\dot{X}$ in Eq. (5.33))

$$3H\dot{\phi} \simeq -V'\,. \tag{5.38}$$

For concreteness and without loss of generality we will assume henceforth that $V' > 0$ and therefore ϕ is decreasing, $\dot{\phi} < 0$. There is a bit more to this equation than meets the eye:

- The second order EOM has become a first order one, which can be straightforwardly integrated (at least formally)
- The righthand side depends only on the shape of the potential, while the lefthand side really knows about the specific solution. This equation is therefore the bridge between Hubble and potential slow-roll parameters[54].

[54] CFU: *Use Eq. (5.38) and Eq. (5.25) to show that* $\epsilon \simeq \epsilon_V$.

- Third, in this approximate equation, $\dot{\phi}$ is fixed once we specify ϕ. We will see that this remarkable simplification is somewhat an accident of having a single field and does not generalize to two or more fields.

Combining the two approximate equations of motion (5.38) and (5.35) one can reduce the problem to solving a non-linear 1st order ordinary differential equation It is often more convenient to re-write the equation in terms of e-foldings N rather than time t using $\partial_N \phi = H\dot{\phi}$. Then we find

$$\partial_N \phi \simeq -\frac{V' M_{\rm Pl}}{3H^2} \simeq -M_{\rm Pl}^2 \frac{V' M_{\rm Pl}}{V} = -M_{\rm Pl}\sqrt{2\epsilon}\,. \tag{5.39}$$

Before solving this equation we notice that inflation ends by definition when $\epsilon > 1$, which is when accelerated expansion gives way to decelerated expansion, see (5.1). When this happens the slow-roll approximation has broken down because ϵ is not small anymore. However, as a rough estimate, we can say that inflation when $\phi = \phi_e$ with

$$\epsilon(\phi_f) \simeq \epsilon_V(\phi_e) = 1 \quad \Rightarrow \quad M_{\rm Pl} V'(\phi_e) = 2V(\phi_e)\,. \tag{5.40}$$

In many simple models inflation ends because we approach a minimum of the potential at ϕ_{min}. For example, for a monomial potential $V = \lambda_p \phi^p$ this would happen at $\phi_e = (p/\sqrt{2})M_{\rm Pl}$. However, it is also possible that the potential stops being slow-roll steep and the inflation fast rolls down for some time before settling in a minimum. For consistency with the late universe and the rate of the current acceleration of the universe, one typically assumes that the energy at the minimum matches the cosmological constant today, i.e. $V(\phi_{min}) \sim (10^{-3}\text{eV})^4$. This is such a tiny energy as compared with the typical scale of inflation, (5.5), that we might as well assume $V(\phi_{min}) = 0$ for all practical purposes and ϵ_V generically blows up as we approach it. Integrating both side of (5.39) from the end of inflation at ϕ_e gives

$$\Delta N(\phi) = \int_{\phi_e}^{\phi} \frac{d\phi}{M_{\rm Pl}} \frac{1}{\sqrt{2\epsilon}} = \int_{\phi}^{\phi_e} \frac{d\phi}{M_{\rm Pl}} \frac{V}{V' M_{\rm Pl}}\,. \tag{5.41}$$

The resulting $\phi(N)$ is the slow-roll solution, which is a good approximation to the exact solution when $\epsilon, \eta \ll 1$. Mountains of papers have been written about the infinitely many possible choices of $V(\phi)$ (see e.g. [40] for an older and [44] for a recent review). To remain agnostic about $V(\phi)$, let us make the very rough approximation that $\sqrt{2\epsilon_V}$ does not vary much for most of the duration of inflation. Then (**??**) gives the relation

$$\frac{\Delta\phi}{M_{\rm Pl}} \sim \Delta N \frac{M_{\rm Pl} V'}{V}\,. \tag{5.42}$$

This tells us that, to achieve a given number of efoldings, say, $\Delta N \sim 50$, flat potentials need a small field excursion $\Delta\phi = \phi_e - \phi_i$, while steep potential need a large field excursion. It customary to divide inflationary potentials into *small field* or *large field* models, depending on wether $\Delta\phi < M_{\rm Pl}$ or $\Delta\phi > M_{\rm Pl}$, respectively. Then (5.42) tells us that potentials that vary on a parametrically subPlanckian scale $\Lambda_\phi \ll M_{\rm Pl}$, defined as $\Lambda_\phi V' \sim V$, lead to superPlanckian field excursions $\Delta\phi \gg M_{\rm Pl}$ and vice versa. There is an ongoing very active and controversial debate as to whether these large field models are allowed in a consistent quantum theory of gravity. As the inflaton oscillated around the minimum of the potential, with ever decreasing amplitude due to the Hubble friction term in (5.23), quantum processes become relevant and the inflaton decays into a hot soup of standard model particles. This process is known as reheating.

Monomial potentials As an example, we discuss one simple class of potentials here, but the reader can play with a different class in Prob. P.5.2. Let's work our the predictions of monomial potentials of the form

$$V = \lambda_p \phi^p \,, \tag{5.43}$$

with λ_p an unknown coupling and p a positive real number. Recalling that inflation ends when $\epsilon_V = 1$ and so $\phi = (p/\sqrt{2})M_{\rm Pl}$, we find the following relation between ϕ and the number of e-foldings to the end of inflation

$$\Delta N(\phi) = \frac{1}{2p}\left(\frac{\phi^2}{M_{\rm Pl}^2} - \frac{p^2}{2}\right) . \tag{5.44}$$

As we will learn, the modes we observe in the large scale structures and the CMB left the Hubble radius some number $\Delta N_{\rm infl}$ of e-foldings before the end of inflation. Estimating this uncertain number given in (5.8) as $\Delta N_{\rm infl} \sim 50$, we evaluate

$$\frac{\phi_{50}}{M_{\rm Pl}} = \sqrt{2p\left(\Delta N_{\rm infl} + \frac{p}{4}\right)} . \tag{5.45}$$

For example, for $p = 2$ this gives $\phi_{50} \sim 14 M_{\rm Pl}$. The slow-roll parameters are found to be

$$\epsilon \sim \epsilon_V = \frac{M_{\rm Pl}^2}{2}\left(\frac{p\lambda\phi^{p-1}}{\lambda\phi^p}\right)^2 = \frac{p^2}{2}\frac{M_{\rm Pl}^2}{\phi^2} \,, \tag{5.46}$$

$$\eta_V = M_{\rm Pl}^2 \frac{p(p-1)\lambda\phi^{p-2}}{\lambda\phi^p} = p(p-1)\frac{M_{\rm Pl}^2}{\phi^2} . \tag{5.47}$$

We see that for $\phi \gg M_{\rm Pl}$ both are indeed small and so the slow-roll approximation is justified. Also, much before the end of inflation, where $\phi \gg M_{\rm Pl}$, using (5.45) we can write

$$\epsilon_V \simeq \frac{p}{4\Delta N} \,, \qquad \eta_V = \frac{p-1}{2\Delta N} . \tag{5.48}$$

Problems for lesson 5

P.5.1 For $p > 0$, consider the simple "chaotic inflation" potential

$$V(\phi) = \lambda_p \phi^p \,, \tag{5.49}$$

(a) What is the mass dimension of λ_p

(b) When are the potential slow-roll parameters ϵ_V, η_V small?

(c) At what ϕ_e does acceleration end (recall $\ddot{a} > 0 \to \epsilon < 1$)?

(d) For $p = 2$ find $N(\phi)$ in slow-roll. What ϕ_i gives $N_e = 50$?

P.5.2 Consider a canonically normalized scalar field ϕ with the potential

$$V = V_0\left[1 + \cos\left(\frac{\phi}{f}\right)\right] \,, \tag{5.50}$$

with V_0 setting the overall vertical scale and the *axion decay constant* f setting the horizontal scale.

(a) What symmetries does this theory enjoy?

(b) Compute ϵ_V and η_V for this potential, as function of ϕ. Notice how they depend on the overall scale V_0

(c) Estimate ϕ_{CMB} corresponding to 60 efoldings before the end of inflation

(d) In what regime of the parameters f and V_0 does this potential become indistinguishable, during the last 60 efoldings of inflation, from the quadratic potential $m^2\phi^2/2$?

P.5.3 Derive the equations of motion or the $P(X, \phi)$ theories.

(a) Derive the equations of motion (5.17) by varying the action $\delta S/\delta\phi$.

(b) Compute the energy-momentum tensor for homogeneous configurations $\phi = \phi(t)$.

(c) From $T_{\mu\nu}$, compute the energy density ρ and the pressure p.

(d) Derive the Friedmann and acceleration equations (5.18) by using the general expression (1.43) and (1.52), and the expression for ρ and p in terms of $P(X, \phi)$ and its derivatives you computed previously.

(e) Specify to $P = P(X)$ and prove that a stationary point X_s where $\partial_X P(X)|_{X_s} = 0$ provides a solution the EoM. This is called the ghost condensate [4]. What spacetime solution emerges?

P.5.4 Around (5.8) we compute the minimum number of efoldings to solve the horizon problem. Compute the lower bound on ΔN_{inf} obtained by requiring to solve the curvature problem, assuming that at the beginning of inflation $\Omega_k \lesssim \mathcal{O}(1)$ (but you are allowed to neglect K in the Friedmann equation).

6 Thermal history: equilibrium

6.1 Thermal history

Before developing the necessary mathematical formulation, we will start with a broad stroke description of the history of the universe, using time and temperature to characterise different important moments. To move between t and T quickly we can use the simple, approximate relation $T_{\rm MeV} \sim 1/\sqrt{t_{\rm sec}}$. A heuristic derivation[55] of this relation goes as follows. Recall that the equation of state parameter radiation is $w = 1/3$. Moreover, we make use of $\rho \propto g_\star T^4$, which we derive around (6.22), where T is the temperature and $g_\star$ the number of relativistic degrees of freedom. In the standard model $g_\star$ ranges from around 100 at high temperature to a few at low temperatures. For this heuristic derivation we will assume it is constant, $g_\star \sim 100$. Then

$$0 = \dot\rho + 4H\rho \propto \left(\dot T + HT\right) . \tag{6.1}$$

So solve this differential equation for $T(t)$ we need $H(t)$, which is given by the Friedmann equation

$$H = \sqrt{\frac{\rho}{3M_{\rm Pl}}} = \sqrt{g_\star \frac{\pi^2}{90}} \frac{T^2}{M_{\rm Pl}} . \tag{6.2}$$

We can then solve (6.1) to find

$$\dot T = -T^3 \frac{\pi}{M_{\rm Pl}} \sqrt{\frac{g_\star}{90}} \tag{6.3}$$

and hence $T(t) \propto 1/\sqrt{t}$. Expressing this relation in convenient units one finds

$$t = 0.25 \text{ sec} \times \left(\frac{\rm MeV}{T}\right)^2 , \qquad T = 0.5 \text{ MeV} \times \left(\frac{\rm sec}{t}\right)^{1/2} . \tag{6.4}$$

These results neglect the variation of $g_\star$ and assumes the Universe is dominated by radiation and so it is valid only much before matter-radiation equality around $z = 3000$ or 60 thousand years. A summary of the conversion is provided in table Tab. 1.

Let us know briefly discuss the most important cosmological events, their time and energy scales in chronological order:

- **Quantum gravity** $T \sim 10^{18}$ GeV, approximately 10^{-43} sec: the perturbative quantum description of GR breaks down and the theory needs an Ultra-Violet (UV) completion. For example, new, unknown degrees of freedom could appear at or before this scale. This happens e.g. in *string theory*, where higher-spin states become dynamical at the string scale[56] $M_s \lesssim M_{\rm Pl}$. Alternatively the theory could become strongly coupled and we don't know what happens. It has been conjectured that GR might possess an UV-fixed point,

[55] A correct derivation would use the conversion of entropy, which we will discuss in Sec. 6.4.

[56] The value of the string scale is of course unknown and depends on the details of the compactification from 10 (or 11) down to 4 dimensions.

where all coupling constants of the theory, including all higher dimension operators, have vanishing beta functions. This line of investigation goes under the name of *asymptotic safety*.

- **Inflation and reheating.** $H \sim 10^{-24}$ GeV - 10^{13} GeV, a conjectured phase of accelerated expansion called cosmological *inflation* seeds the primordial perturbation that later will give rise to the structure in the universe and eventually to us. The energy scale of this process is one of the most uncertain scales in physics. During inflation, the universe is cold and empty, the abundance of any standard model species is exponentially suppressed in time by the fast expansion $a \sim e^{Ht}$, with H approximately constant. The universe expands by at least a factor of approximately[57] $a_f/a_i \sim e^{60} \sim 10^{26}$. Inflation is driven by some degrees of freedom collectively known as the *inflaton sector*, or sometimes simpley the inflaton even though there could be more than one. At the end of inflation the inflaton breaks up into particles, which in turns decay into standard model fields in a process called *reheating*. In the simplest and most standard paradigm, this final state consists of a hot thermalized soup of standard model particles with $T >$TeV. The *hot big bang* starts here.

- **Baryogenesis.** At $T >$100 GeV, an asymmetry in baryon number is created by some, yet unknown, non-equilibrium, P- and CP-violating process [55] called *baryogenesis*. As all quarks annihilate with anti-quarks, and only a part in a million of the baryonic matter in the universe survives. This will eventually form all atoms in the universe.

- **Electroweak symmetry breaking.** $T \sim 100$ GeV - 10^3 GeV, the *electroweak symmetry* of the standard model $SU(2) \times U_Y(1)$ is broken via the Brout-Englert-Higgs mechanism down to the abelian $U(1)$ gauge symmetry that we call electromagnetism. The details of this phase transition depend crucially on the properties of the Higgs particle and of the spectrum of the standard model, which are being currently probed at particle accelerators such as the Large Hadron Collider at CERN. All standard model fermions, namely quarks and leptons, as well as the $W^{\pm}$ and Z^0 vector bosons acquire a mass proportional to the vacuum expectation value of the Higgs field.

- **QCD phase transition.** At $T \sim 200$ MeV the free quarks and gluons become confined as the coupling of the strong interactions becomes of order one. Because of its non-perturbative nature, the details of this *QCD phase transition* leading to confinement are still not fully understood. As the temperature decreases below the mass of the lightest mesons, namely the pions $\pi^{\pm,0}$ whose mass is protected by the approximate global isospin symmetry, all quarks and gluons in the universe become confined inside protons and neutrons, which obey a thermal distribution.

- **Neutrino decoupling.** At $T \sim 1 - 3$ MeV, or $t \sim 0.2$ sec, neutrinos fall out of equilibrium as their weak interaction rate becomes smaller than the expansion of the universe. Different neutrino flavor, $\nu_{e,\mu,\tau}$ decouple at slightly different energies. From this moment onward, neutrinos couple only gravitationally and mostly free stream across the universe.

[57] The exact number of factors of e, namely $N \equiv \ln(a_f/a_i)$, a.k.a. *efoldings*, is not known. Many inflationary models have $40 < N < 60$, while data constraints $N > 20$.

- **Neutron freeze-out.** At $T \sim 1$ MeV the neutrons fall out of thermal equilibrium and their abundance freezes out, up to a small decaying rate which, on the time scale of the problem, produces only an order 10% effect. The ratio of protons to neutrons in the universe is approximately fixed by this process.

- **Electron-positron annihilation.** At $T \sim 0.5$ MeV, or $t \sim 5$ sec electron-positrons annihilation takes place. As the temperature drops below the electron mass 0.5 MeV, the process of electron-positron production becomes very rare and all positrons annihilate with electrons. As we observe an electrically neutral universe, a number of electron survive equal to the number of protons. As discussed around Eq. (6.62), this process releases energy into the photons, which therefore become hotter than the neutrinos (which had decoupled early).

- **Big bang nucleosynthesis.** At $T \sim 0.07$ MeV or $t \sim 3$ minutes and $z = 10^{10}$, protons and neutrons combine to form Deuterium, the isotope of Hydrogen with one proton and one neutron. It in turn converts almost immediately into Helium-4. The capture of neutrons, whose lifetime would be around 15 minutes, to form nuclei prevents them from decaying further. The primordial abundance of atoms is determined in this process, which is known as big bang nucleosynthesis (BBN). For the lightest elements of the periodic table, BBN abundance gets modified only marginally by subsequent astrophysical processes. The prediction of the abundance of light atoms is one of the greatest successes of the big bang theory.

- **Black-body time.** At $T \sim 0.5$ keV, or $t = 2$ months, $z = 2 \times 10^6$, the number of photons with energy of the order of T and above becomes effectively frozen because all active interactions, such as Compton scattering, conserve photon number. This is the black-body time, after which any process involving photons can destroy the black-body spectrum of the photons, which we eventually measure in the Cosmic Microwave Background radiation (CMB).

- **Matter-radiation equality.** At $T \sim 1$eV, or $z = 3300$, the energy density of all matter equals that of radiation, which is known as matter-radiation equality. Here matter includes dark, which is 6 parts out of 7, and baryonic matter, which constitutes the remaining 1 part in 7. Radiation is made of photons, around 60%, and neutrinos, the remaining 40%. This moment it in time is important because dark matter inhomogeneities start growing at this point to eventually form structures.

- **Recombination.** At $T \sim 0.3$ eV, around 370 ky, $z \simeq 1100$ recombination of electron and protons to form neutral Hydrogen takes place. Earlier, at $z = 1400$, we have the recombination of Helium, which captures two electron per nucleus; this is a smaller effect and can be neglected for rough estimates. The fraction of free charged particles, namely e^- and p^+, decays very fast and quickly the photon cross section for Compton scattering becomes negligible. The universe becomes transparent. Most of the photons travel freely in every direction. It is these photons that we detect as CMB.

- **Re-ionization**. At $z \sim 10$ most of the Hydrogen in the universe becomes ionized again as stars and galaxies become abundant. The detailed of this process, known as *reionization*, are still very uncertain and are expected to be clarified by ongoing and near

z	Size	Temperature	Age	Comov. Dist.	Part. Horizon	Energy
0	14.3 Gpc	0.000234 eV	13.7 Gy	0.	14.3 Gpc	Λ
0.1	13.0 Gpc	0.000258 eV	12.4 Gy	414 Mpc	13.9 Gpc	Λ
0.39	10.3 Gpc	0.000326 eV	9.48 Gy	1.51 Gpc	12.8 Gpc	$\Lambda = \Omega_m$
1.	7.15 Gpc	0.000469 eV	5.92 Gy	3.32 Gpc	10.9 Gpc	Ω_m
3	3.57 Gpc	0.000937 eV	2.19 Gy	6.46 Gpc	7.81 Gpc	Ω_m
6	2.04 Gpc	0.00164 eV	947 My	8.42 Gpc	5.85 Gpc	Ω_m
10	1.3 Gpc	0.00258 eV	480 My	9.66 Gpc	4.61 Gpc	Ω_m
20	681 Mpc	0.00492 eV	181 My	11.0 Gpc	3.27 Gpc	Ω_m
50	280 Mpc	0.0120 eV	47.4 My	12.3 Gpc	2.00 Gpc	Ω_m
100	142 Mpc	0.0237 eV	16.8 My	12.9 Gpc	1.35Gpc	Ω_m
1100	13.0 Mpc	0.258 eV	369 ky	14.0 Gpc	280 Mpc	Ω_m
3200	4.47 Mpc	0.750 eV	56.9 ky	14.1 Gpc	119 Mpc	$\Omega_m = \Omega_r$
5×10^4	286 kpc	11.7 eV	292 y	14.3 Gpc	9.01 Mpc	Ω_r
$2. \times 10^6$	7.15 kpc	468.7 eV	68.0 days	14.3 Gpc	0.229 Mpc	Ω_r

Table 1: Numerical conversion among various measures of time in for the best-fit value ΛCDM.

future observations especially by the satellite experiment James Webb Space Telescope. A small fraction of CMB photons, about 10%, scatters again off the now very diluted free electrons.

- **Accelerated expansion.** Around $z \sim 0.3$, or 9 Gy, the matter energy density becomes comparable to that of dark energy and the universe enters a phase of accelerated expansion. Structure formation comes to a stop because the expansion of the universe wins over gravitational collapse.
- **Present.** At $z \sim 0$, around 14 Gy, these lecture notes are written.

6.2 Relativistic kinetic theory

To provide a quantitative description of the thermal history of the universe we have to recall some basic aspects of statistical physics and relativistic kinetic theory. In a many-particle system it is convenient to adopt a probabilistic description of the position and momentum of an average particle. The phase space density $f(x,p,t)$ for particles of mass m is defined as the infinitesimal probability $d\text{Prob}$ of finding a particle at comoving position $\mathbf{x}$ in a physical volume $(ad\mathbf{x})^3$ with physical momentum $\mathbf{q}$ at time t

$$d\text{Prob} = f(\mathbf{x},\mathbf{q},t)a^3 d^3x d^3q\,. \tag{6.5}$$

Here we found it convenient to parameterize the comoving four-momentum of an on-shell particle of mass m in terms of the physical three momentum $\mathbf{q}$ using

$$P^\mu = (E, P^i) = \left(\sqrt{m^2+q^2}, \frac{\mathbf{q}}{a}\right)\,, \tag{6.6}$$

where $q = |\mathbf{q}|$. The energy-momentum tensor $T_{\mu\nu}$ and number density n for a certain species of particles are then

$$T^{\mu\nu}(\mathbf{x},t) = g\int \frac{d^3\mathbf{q}}{(2\pi)^3}\,\frac{P^\mu P^\nu}{E} f(\mathbf{x},\mathbf{q},t)\,, \tag{6.7}$$

$$n(\mathbf{x},t) = g\int \frac{d^3\mathbf{q}}{(2\pi)^3} f(\mathbf{x},\mathbf{q},t)\,, \tag{6.8}$$

where g is the degeneracy of the one-particle state, equivalently the number of propagating degrees of freedom. For example, $g = 2$ for massless vectors such as the photon or massless tensors such as the graviton, which can have helicities ± 1 and ± 2 respectively; $g = 2$ for a Dirac fermion such as the electron e^-, the positron e^+ or the proton p^+, with helicities $\pm 1/2$; and $g = 1$ for a Weyl or Majorana fermion such as the neutrino and its anti-particle[58].
It is convenient to adapt the very general integrals in (6.7) to the case of most relevance in cosmology. Let us then consider particles that are in equilibrium, and therefore obey [59] Bose-Einstein or Fermi-Dirac statistics[60]

$$f(\mathbf{x},\mathbf{q},t) = f_{BE,FD}(\mathbf{x},|\mathbf{q}|,t) = \frac{1}{e^{(E-\mu)/T} \mp 1}\,, \tag{6.9}$$

where $E = P^0 = \sqrt{m^2+q^2}$ and the spacetime dependence appears in the *chemical potential* $\mu = \mu(\mathbf{x},t)$ and the temperature $T = T(\mathbf{x},t)$, which in natural units $k_B = 1$ have both dimension of energy.
To help with intuition, let's recall the definition of the chemical potential. It is given by the derivative of the entropy with respect to the particle number, at fixed volume and energy:

$$\mu = -T\left(\frac{\partial S}{\partial N}\right)_{U,V}\,. \tag{6.10}$$

The differential change in entropy of a system then reads

$$dS = \frac{1}{T}dU + \frac{P}{T}dV - \frac{\mu}{T}\,dN\,. \tag{6.11}$$

This expression highlights a close analogy between inverse temperature $1/T$ and rescaled chemical potential $-\mu/T$. At lower temperature the entropy gain per unit energy is larger than at a higher temperature because $1/T$ is large. The flow of energy U from a hot region to a colder one hence increases the total entropy and is therefore statistically favored, as expected from the second law of thermodynamics. In a similar way, when the chemical potential differs between two regions, particles will move from a region where $-\mu/T$ is smaller to another where

[58] It is not known whether the neutrino is a Weyl or a Majorana particle. Either way the final counting is the same: the 4 real components for a Majorana spinor can be written in terms of the 2 complex components for a Weyl spinor. A Weyl spinor is a chiral particle (e.g. left-handed for neutrinos), with an antiparticle of opposite chirality (the right-handed anti-neutrino). A Majorana particle instead has both chiralities and is its own anti-particle.

[59] Note that both distributions reduce to the Maxwell-Boltzmann distribution $f_{MB} = e^{-(P^0-\mu)/T}$ when the occupation numbers are small, namely in the limit $e^{(P^0-\mu)/T} \gg 1$. This limit arises when the density is low and the chemical potential is very negative, $\mu \ll -T$.

[60] CFU: *How do you get the right sign in the denominator of $f_{BE,FD}$? Remember the exclusion principle for Fermions which implies $f_{FD} < 1$.*

it is larger so that the total entropy increases. The same argument applies for reactions. For simplicity imagine a reaction from particles of species 1 and 2 to particles of species 3 and 4. The reaction will proceed in the forward direction if the total entropy increases, which happens as long as there is a difference in chemical potential. Chemical equilibrium is reached when the chemical potentials of the reactants and products balance, i.e.

$$\mu_1 + \mu_2 = \mu_3 + \mu_4 \,. \tag{6.12}$$

We will later use this equation to relate the chemical potentials of different standard model particles during the hot big bang, see (6.39).
In the following, we will make use of the integrated version of the first law in (6.11), which is sometimes known as Euler's equation[61]

$$TS = E + pV + \mu N \,. \tag{6.13}$$

Dividing by the volume and defining the energy, number and entropy densities $\rho = E/V$, $n = N/V$ and $s = S/V$ we find

$$s = \frac{\rho + p + \mu n}{T} \,. \tag{6.14}$$

By homogeneity and isotropy, the only non-vanishing components of the energy momentum tensor are $T^0_{\ 0} = -\rho$ and $T^i_{\ i} = 3p$. Using spherical coordinates, the angular integrations in (6.7) simply give a factor of 4π leaving

$$\rho = \frac{g}{2\pi^2} \int dq\, q^2 \, \frac{E}{e^{(E-\mu)/T} \mp 1} \,, \tag{6.15}$$

$$p = \frac{g}{2\pi^2} \int dq\, q^2 \, \frac{q^2}{3E} \frac{1}{e^{(E-\mu)/T} \mp 1} \,, \tag{6.16}$$

$$n = \frac{g}{2\pi^2} \int dq\, q^2 \frac{1}{e^{(E-\mu)/T} \mp 1} \,, \tag{6.17}$$

for bosons and fermions respectively, with $E = \sqrt{m^2 + q^2}$.

Relativistic particles For $T \gg m$, these integrals are mostly supported around $q \simeq 3T$ and so at high temperature we can approximate $E(q) = \sqrt{m^2 + q^2} \simeq q$ up to corrections of order (m/T). At this order $p^2/3E = E/3$ and so $\rho = 3p$. Performing the integrals [62] above one finds

$$\rho = 3p = g \frac{3}{\pi^2} T^4 \left[\pm \mathrm{Li}_4(\pm e^{\mu/T}) \right] \,, \tag{6.20}$$

$$n = g \frac{1}{\pi^2} T^3 \left[\pm \mathrm{Li}_3(\pm e^{\mu/T}) \right] \,, \tag{6.21}$$

[61] Yet another confirmation of Stigler's law of eponymy.
[62] Here we have used the master integral

$$\int_0^\infty dy \, \frac{y^n}{e^{y-z} - \eta} = \frac{1}{\eta} \Gamma(n+1) \mathrm{Li}_{n+1}(e^z \eta) \tag{6.18}$$

where the polylogarithm is the generalisation of the logarithm in the sense that $\mathrm{Li}_1(z) = -\log(1-z)$ and

$$\mathrm{Li}_{n+1}(z) = \int_0^z \frac{\mathrm{Li}_n(z')}{z'} dz' \,. \tag{6.19}$$

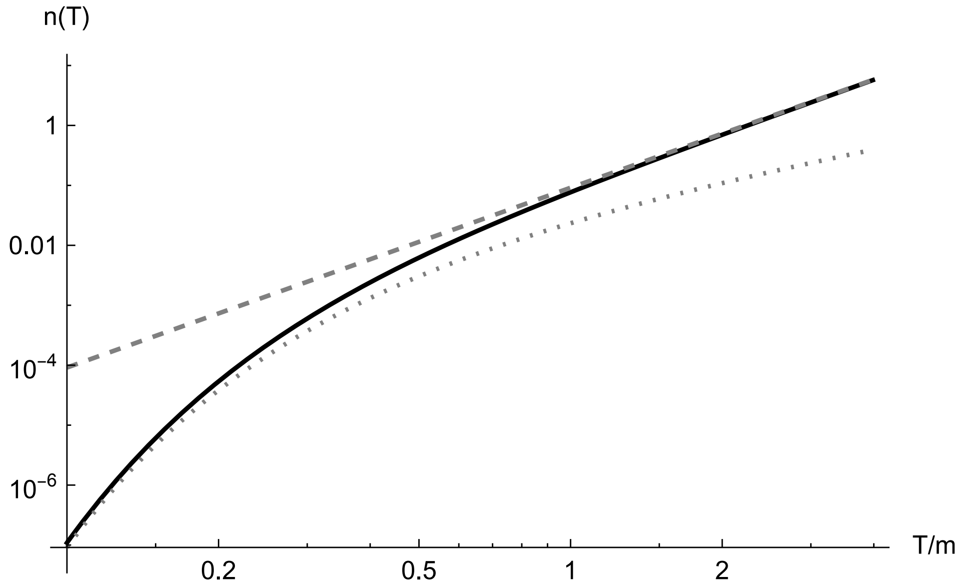


Figure 13: The comparison of the relativistic (gray dashed line) and non-relativistic (gray dotted line) approximations to the exact expression (black line) for the number density n as function of T/m and assuming $\mu = 0$.

for bosons and fermions respectively, where $\mathrm{Li}_n(z)$ is the polylogarithm of z at order n. This result explains why relativistic particles are often called *radiation*, since $\rho = 3p$ is the equation of state of radiation, $w = 1/3$, which we introduced in Eq. (1.33).
The chemical potentials are small for all known particles at almost all times, so we can simplify these expressions in the limit $\mu \ll T$. The polylogarithms can be evaluated as

$$\mathrm{Li}_4(1) = \pi^4/90\,, \quad \mathrm{Li}_3(1) = \zeta(3) \sim 1.2$$

and one finds

$$\rho = 3p = g\frac{\pi^2}{30}T^4 \begin{cases} 1 & \text{(relativistic bosons)} \\ \frac{7}{8} & \text{(relativistic fermions)} \end{cases}\,, \tag{6.22}$$

as well as

$$n = g\frac{\zeta(3)}{\pi^2}T^3 \begin{cases} 1 & \text{(relativistic bosons)} \\ \frac{3}{4} & \text{(relativistic fermions)} \end{cases}\,, \tag{6.23}$$

This approximation is compare to the exact integral in Fig. 13. We can also compute the entropy density (always neglecting μ)

$$s = \frac{\rho + p}{T} = g\frac{2\pi^2}{45}T^3 \begin{cases} 1 & \text{(relativistic bosons)} \\ \frac{7}{8} & \text{(relativistic fermions)} \end{cases}\,. \tag{6.24}$$

Non-relativistic particles In the opposite limit, at low temperatures $m - \mu \gg T$ both quantum statistics reduce to the Boltzmann distribution since

$$e^{(\sqrt{m^2+q^2}-\mu)/T} > e^{(m-\mu)/T} \gg 1\,. \tag{6.25}$$

Now the integral is mostly supported around $q \simeq \sqrt{Tm} \ll m$. If we also assume $m \gg T$, we can approximate $\sqrt{m^2+q^2} \simeq m + q^2/(2m)$ everywhere, up to correction of order $T/m \ll 1$. Then the integrals can be done analytically and the result is

$$n = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{(\mu-m)/T}\,, \tag{6.26}$$

$$\rho = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{(\mu-m)/T}\left(m + \frac{3}{2}T\right) = n\left(m + \frac{3}{2}T\right)\,, \tag{6.27}$$

$$p = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{(\mu-m)/T}T = nT\,. \tag{6.28}$$

$$s = \frac{\rho+p}{T} = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{(\mu-m)/T}\left(\frac{m}{T} + \frac{5}{2}\right)\,. \tag{6.29}$$

Note that $p = nT$ is nothing but the well-known *ideal gas law*, $pV = Nk_BT$ written in natural units $k_B = 1$ in terms of the number density $n = N/V$. These expressions justify the name *non-relativistic matter* for particles with $T \ll m$, since

$$\rho \simeq mn \quad \& \quad p = nT \quad \Rightarrow \quad p \ll \rho\,, \tag{6.30}$$

which is the equation of state of non-relativistic matter, $w \simeq 0$. In Fig. **??** we show compare the relativistic and non-relativistic approximations for $n(T)$ as function of T/m for $\mu = 0$. The optimal approximation is given by switching from the non-relativist to the relativist approximation around $T/m \simeq 0.6$, where each approximation is about 50% higher or lower than the exact result.

An important observation is that, if relativistic and non-relativistic particles are in thermal equilibrium, i.e. at the same temperature, and $\mu \ll T$, then

$$\frac{\rho_{\text{non-rel}}}{\rho_{\text{rel}}} \propto e^{-m/T}\left(\frac{T}{m}\right)^{5/2} \ll 1\,. \tag{6.31}$$

In words, the energy density of non-relativistic particles is exponentially suppressed as compared with that of relativistic particles at the same temperature. The same is true for the pressure and entropy density. We conclude that in thermodynamical equilibrium, *particles become irrelevant for the total energy, pressure and entropy budget as soon as they become non-relativistic.*

The number of relativistic degrees of freedom Since relativistic species dominate the energy budget of the universe in thermal equilibrium, it convenient to have a simple expression for the radiation energy density. Let us introduce the effective number of bosonic degrees of freedom $g_\star$ defined as

$$\rho = g_\star \frac{\pi^2}{30} T^4 \,, \quad \text{with} \quad g_\star \equiv g_{\rm Bos} + \frac{7}{8} g_{\rm Fer} \,. \tag{6.32}$$

The effective number of relativistic degrees of freedom can be split into two parts,

$$g_\star = g_\star^{\rm th} + g_\star^{\rm dec},$$

where $g_\star^{\rm th}$ accounts for species in thermal contact with the photons, while $g_\star^{\rm dec}$ describes those that have decoupled. Particles in equilibrium with photons must all have the same temperature T, and if they are relativistic $T \gg m_i$. Then their contribution is

$$g_\star^{\rm th}(T) = \sum_{i \in {\rm Bos}} g_i + \frac{7}{8} \sum_{i \in {\rm Fer}} g_i \ , \tag{6.33}$$

with sums over bosonic and fermionic degrees of freedom. If the temperature of a species i falls below its mass, $T_i < m_i$, the species becomes non-relativistic and is no longer included. We call this a threshold. Away from thresholds, $g_\star^{\rm th}(T)$ is essentially constant in T. We have to also account for relativistic species that are no longer in equilibrium, $T_i \neq T \gg m_i$. Their energy density is proportional to T_i^4 and so, if we use the definition in (14) we must rescale $g_\star^{\rm dec}$ by $(T_i/T)^4$:

$$g_\star^{\rm dec}(T) = \sum_{i=b} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{i=f} g_i \left(\frac{T_i}{T}\right)^4 \ . \tag{6.34}$$

This case arises when particles decouple, an effect particularly relevant for neutrinos.
Given the standard model spectrum in Table 2, one can compute the effective number of relativistic degrees of freedom as a function of temperature. At very high temperatures, $T \gtrsim 100$ GeV, all standard model fields are relativistic and contribute:

$g_{\rm Bos} = 28$ from photons (2), $W^\pm$ and Z^0 (3×3), gluons (8×2), and the Higgs (1)
$g_{\rm Fer} = 90$ from quarks (6×12), charged leptons (3×4), and neutrinos (3×2)

leading to

$$g_\star = g_{\rm Bos} + \tfrac{7}{8}\, g_{\rm Fer} = 106.75 \ . \tag{6.35}$$

As the Universe cools, species become non-relativistic and annihilate, so $g_\star(T)$ decreases step by step. For instance, near $T \sim 10$ MeV one finds

$$g_\star = 2 + \tfrac{7}{8}(2 \times 2 + 3 \times 2) = 10.75,$$

coming from photons (2 polarizations), electrons and positrons (2 spins each), and three neutrino families with their antiparticles. Subsequent evolution involves the neutrinos decoupling from electrons and photons, followed by the crossing of the electron threshold, which is also known as electron-positron annihilation. To analyze these processes it is useful to first establish entropy conservation.

Table 2: Particle content of the Standard Model. Credit: D. Baumann.

type		mass	spin	g
quarks	$t, \bar{t}$	173 GeV	$\frac{1}{2}$	$2 \cdot 2 \cdot 3 = 12$
	$b, \bar{b}$	4 GeV		
	$c, \bar{c}$	1 GeV		
	$s, \bar{s}$	100 MeV		
	$d, \bar{d}$	5 MeV		
	$u, \bar{u}$	2 MeV		
gluons	g_i	0	1	$8 \cdot 2 = 16$
leptons	$\tau^\pm$	1777 MeV	$\frac{1}{2}$	$2 \cdot 2 = 4$
	$\mu^\pm$	106 MeV		
	$e^\pm$	511 keV		
	$\nu_\tau, \bar{\nu}_\tau$	< 0.6 eV	$\frac{1}{2}$	$2 \cdot 1 = 2$
	$\nu_\mu, \bar{\nu}_\mu$	< 0.6 eV		
	$\nu_e, \bar{\nu}_e$	< 0.6 eV		
gauge bosons	W^+	80 GeV	1	3
	W^-	80 GeV		
	Z^0	91 GeV		
	γ	0		2
Higgs boson	H^0	125 GeV	0	1

Table 3: The degrees of freedom of the standard model and their corresponding mass and spin.

6.3 Chemical potential and chemical equilibrium

In thermal equilibrium, the state of a particle species is fully characterised by two macroscopic quantities: its temperature $T(t)$ and its chemical potential $\mu(t)$. These determine, respectively, how energy and particle number are distributed among the accessible microstates. It is therefore important to distinguish between *kinetic equilibrium*, which fixes the shape of the momentum distribution, and *chemical equilibrium*, which fixes the relative abundances of different species.

Kinetic vs. chemical equilibrium. Kinetic equilibrium is maintained by rapid interactions that redistribute energy and momentum without changing the total number of particles of a given species. Elastic scatterings such as

$$e^- + e^- \leftrightarrow e^- + e^-, \qquad \nu_e + e^- \leftrightarrow \nu_e + e^-, \tag{6.36}$$

are typical examples. These processes quickly enforce a Bose–Einstein or Fermi–Dirac distribution

$$f(p,t) = \frac{1}{e^{(E-\mu)/T} \mp 1}, \tag{6.37}$$

with a common temperature for all coupled species, even if the chemical potentials μ_a remain arbitrary. The existence of such fast scatterings allows one to speak of a well-defined local temperature $T(t)$ long after true chemical equilibrium has been lost.

Chemical equilibrium, instead, requires processes that *change the identity or number* of particles, such as annihilations or decays. For a generic reaction $1 + 2 \leftrightarrow 3 + 4$, the condition of chemical equilibrium is

$$\mu_1 + \mu_2 = \mu_3 + \mu_4 \,. \tag{6.38}$$

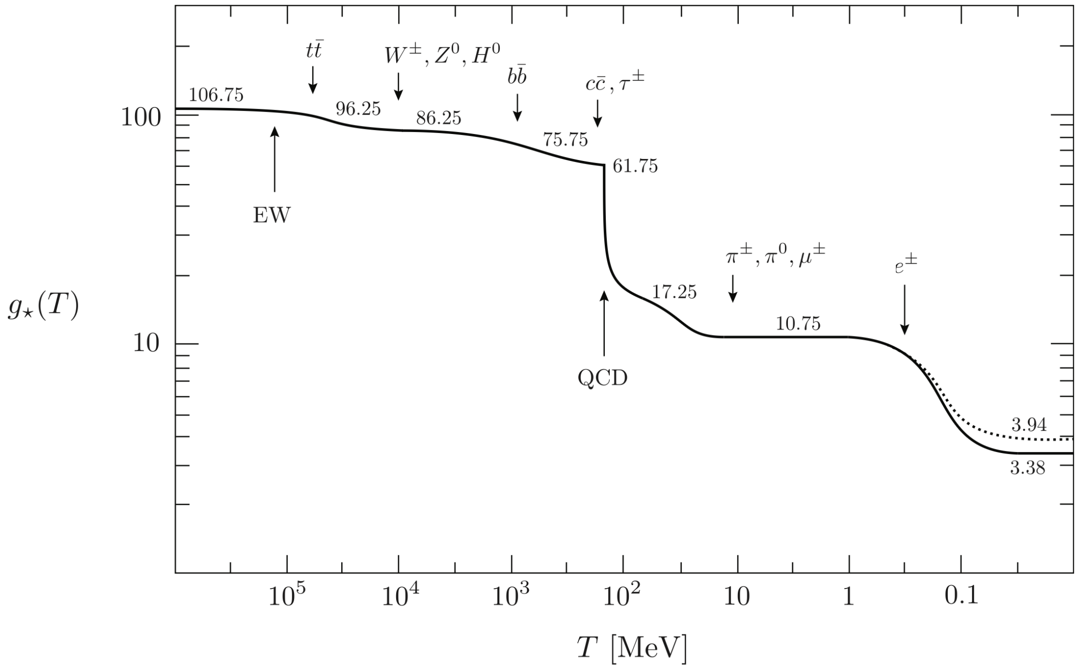


Figure 14: The number of relativistic degrees of freedom $g_\star(T)$ as function of temperature T for the standard model. Figure from D. Baumann.

When this relation holds for all efficient reactions in the plasma, the system has reached full thermodynamic equilibrium. When some reactions becomes inefficient we say it “freeze out”, the corresponding relations among chemical potentials cease to hold, and the associated particle numbers become conserved. We will see how to treat this out-of-equilibrium situations in the next section.

Example: electrons, positrons and neutrinos. Before neutrino decoupling, electrons, positrons and neutrinos are kept in *kinetic* equilibrium by frequent weak and electromagnetic scatterings,

$$\nu_e + e^- \leftrightarrow \nu_e + e^-, \qquad e^- + \gamma \leftrightarrow e^- + \gamma, \tag{6.39}$$

which redistribute energy but do not change particle type. In addition, number-changing processes such as

$$e^+ + e^- \leftrightarrow \nu_i + \bar{\nu}_i, \qquad e^+ + e^- \leftrightarrow \gamma + \gamma, \tag{6.40}$$

maintain *chemical* equilibrium among these species down to temperatures of order $T \sim 1$ MeV. From these reactions one infers the relations among chemical potentials:

$$e^+ + e^- \leftrightarrow \gamma + \gamma \quad \Rightarrow \quad \mu_{e^-} + \mu_{e^+} = 2\mu_\gamma = 0, \tag{6.41}$$

$$\Rightarrow \quad \mu_{e^-} = -\,\mu_{e^+} \equiv \mu_e\,. \tag{6.42}$$

Similarly, for the weak process $e^+ + e^- \leftrightarrow \nu_i + \bar{\nu}_i$,

$$\mu_{e^-} + \mu_{e^+} = \mu_{\nu_i} + \mu_{\bar{\nu}_i} = 0 \quad \Rightarrow \quad \mu_{\bar{\nu}_i} = -\,\mu_{\nu_i}. \tag{6.43}$$

Thus electrons and positrons (and likewise neutrinos and antineutrinos) have opposite chemical potentials in equilibrium.

Chemical potentials from conserved charges. The only exactly conserved quantities of relevance for us are the total electric charge and the baryon and lepton numbers. We observe a globally neutral Universe. Well after baryogenesis, there are no more anti-protons in the universe and so charge neutrality of the Universe implies

$$n_{e^-} - n_{e^+} - n_p \simeq 0, \tag{6.44}$$

At high temperatures, $T \gg m_e$, the densities $n_{e^\pm}$ are obtained from the relativistic Fermi–Dirac integrals given in (6.21),

$$n_{e^-} - n_{e^+} = \frac{g_e T^3}{\pi^2}\left[\mathrm{Li}_3(-e^{\mu_e/T}) - \mathrm{Li}_3(-e^{-\mu_e/T})\right], \tag{6.45}$$

Since we will soon find $|\mu_e| \ll T$, one can expand

$$n_{e^-} - n_{e^+} \simeq \frac{g_e}{3}\,\mu_e T^2. \tag{6.46}$$

To compute μ_e we need to specify the baryon density. A convenient way to express it is through the dimensionless *baryon-to-photon ratio*

$$\eta \equiv \frac{n_b}{n_\gamma}, \tag{6.47}$$

which measures the excess of baryons over antibaryons per photon in the cosmic plasma. After baryogenesis, both the total baryon number and the total entropy are conserved to excellent approximation, while photons dominate the entropy of the Universe. Since the photon number density scales as $n_\gamma \propto T^3$ and the baryon number as $n_b \propto a^{-3}$, their ratio remains approximately constant during the adiabatic expansion. Observations of the cosmic microwave background and of primordial nucleosynthesis give

$$\eta \simeq (6.1 \pm 0.1) \times 10^{-10}. \tag{6.48}$$

Equating (6.46) to the proton density $n_p \simeq n_b = \eta\, n_\gamma$ gives

$$\mu_e \simeq \frac{3 n_b}{g_e T^2} \sim 10^{-9}\, T, \tag{6.49}$$

The electron and positron chemical potentials are therefore extremely small compared with the temperature as long as they are relativistic.
and can safely be neglected in most cosmological applications.
A similar argument applies to neutrinos. If there exists a net lepton number $n_{L_i} = n_{\nu_i} - n_{\bar{\nu}_i}$, one finds

$$n_{\nu_i} - n_{\bar{\nu}_i} = \frac{g_\nu T_\nu^3}{6}\left(\frac{\mu_{\nu_i}}{T_\nu} + \frac{1}{\pi^2}\left(\frac{\mu_{\nu_i}}{T_\nu}\right)^3 + \cdots\right). \tag{6.50}$$

Given the observational bound $|n_{L_i}/n_\gamma| \lesssim 10^{-2}$, this implies

$$|\mu_{\nu_i}|/T_\nu \lesssim 10^{-2}, \tag{6.51}$$

confirming that neutrino chemical potentials are also negligible for the thermal history.
In summary, throughout most of the thermal history we can safely take $\mu_e = \mu_\nu = 0$, while retaining these parameters as bookkeeping devices when discussing small conserved asymmetries.

6.4 Time evolution and conservation of Entropy

Under the assumption of equilibrium, we managed to describe a gas of particles of different species and arbitrary mass in terms of one temperature T and a chemical potential μ for each species. Assuming homogeneity and isotropy, these can only depend on time, $T = T(t)$ and $\mu = \mu(t)$. What dynamical equations should we solve to determine their time evolution? In the next section we will see that it is the Boltzmann equation that determines the evolution both in and out of equilibrium and we will study various examples. However, under certain assumptions EP: *to be continued...*
In cosmology, energy conservation is complicated by the lack of time-translation invariance. A more useful quantity to track is the entropy, which in the early universe is very nearly conserved. Since photons vastly outnumber baryons, the entropy is dominated by the radiation bath, which can be well approximated as being in equilibrium.
Let's evaluate the time derivative of the total entropy $S = a^3 s$ in some volume a^3 using (6.14) and neglecting the chemical potential,

$$\frac{1}{a^3}\frac{d(a^3 s)}{dt} = 3Hs + \frac{d}{dt}\left(\frac{\rho + p}{T}\right) = 3Hs + \frac{\dot\rho + \dot p}{T} - \frac{\dot T(\rho + p)}{T^2}\,. \tag{6.52}$$

For $\dot p$ we can go back to its expression in terms of a momentum integral and write

$$\dot p = \frac{g}{2\pi^2}\int dq q^2 \frac{q^2}{3E}\frac{df}{dt} = -\frac{\dot T}{T}\frac{g}{2\pi^2}\int dq q^2 \frac{q^2}{3}\frac{df}{dE}\,, \tag{6.53}$$

where we used that the only time dependence of f in equilibrium comes from $T = T(t)$ and $f = f(E/T)$. Using $E = \sqrt{q^2 + m^2}$ we can change integration variable from dq to dE, integrate by part df/dE and finally go back to the dq variables to find

$$\dot p = -\frac{\dot T}{T}\frac{g}{2\pi^2}\int dq q^2 \left(\frac{q^2}{3E} + E\right) f = -\frac{\dot T}{T}(\rho + p)\,. \tag{6.54}$$

Plugging this and $\dot\rho = -3H(\rho + p)$ from the continuity equation into (6.52) we find that entropy is conserved in equilibrium, $\dot S = 0$ and so $S = a^3 s$ is constant. It will be useful to introduce some notation.
Recall that entropy density s is defined as $s = S/V$. Adapting (6.14) to multiple relativistic species one can write

$$s = \sum_i \frac{\rho_i + P_i}{T_i} \equiv \frac{2\pi^2}{45} g_{\star S}(T)\, T^3\,, \tag{6.55}$$

where the effective entropy degrees of freedom are

$$g_{\star S}(T) = g_{\star S}^{\rm th}(T) + g_{\star S}^{\rm dec}(T)\,. \tag{6.56}$$

For particles in equilibrium, $g_{\star S}^{\rm th}(T) = g_\star^{\rm th}(T)$. For decoupled particles, since $s_i \propto T_i^3$, one has

$$g_{\star S}^{\rm dec}(T) = \sum_{i=b} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8}\sum_{i=f} g_i \left(\frac{T_i}{T}\right)^3 \neq g_\star^{\rm dec}(T) \ . \tag{6.57}$$

Entropy conservation then implies

$$g_{\star S}(T)\, T^3 a^3 = \text{const}, \qquad \text{or equivalently} \qquad T \propto g_{\star S}^{-1/3} a^{-1}. \tag{6.58}$$

Away from thresholds $g_{\star S}$ is approximately constant and $T \propto a^{-1}$. The extra factor $g_{\star S}^{-1/3}$ reflects the reheating of the remaining relativistic plasma whenever a species becomes non-relativistic and transfers its entropy to the rest. This effect will be illustrated in the following section.

6.5 Neutrinos

Let's apply these ideas to neutrinos. Three periods characterize the evolution of cosmological neutrinos:

- Neutrinos are in thermal equilibrium with SM particles at energies around a few MeV[63]. Neutrinos are very relativistic (MeV$\gg$ 0.17eV) and obey a Fermi-Dirac distribution with $\mu = 0$ and massless dispersion relation.

- Neutrinos decouple before electron-positron annihilation. As long as neutrinos are relativistic ($z \gg 500$), the neutrino temperature is $aT(a) = a_{\rm dec} T_{\rm dec}$.

- Neutrinos became non-relativistic at late times ($z < 500$) and start clustering.

We compute the temperature of neutrinos by relating it to that of photons, namely $T_{\rm CMB}$. An order one effect is the extra energy that photons receive after electron-positron annihilation ($e^+ + e^- \to \gamma + \gamma$), which the neutrinos do not receive because they are already decoupled. Covariant conservation of entropy[64] in an FLRW universe implies $\partial_t(a^3 s) = 0$. Before e^+-e^- annihilation and neutrino decoupling, the total entropy is dominated by relativistic species, as described by (6.55). Hence

$$\frac{45}{2\pi^2} a_1^3 s_1 = (a_1 T_1)^3 g_{\star S} = (a_1 T_1)^3 \left[2 + \frac{7}{8}\, 2 \times (1+1+3)\right] = (a_1 T_1)^3 \frac{43}{4} \,, \tag{6.59}$$

where the bosons are just the two polarizations of the photon, and the fermions are the two helicities of e^-, e^+ and of the three neutrinos[65]. Then neutrinos decouple and their temperature

[63] Different species decouple at slightly different times. Neglecting mass oscillations, one finds $T(\nu_e) \simeq 2.4$MeV and $T(\nu_{\mu\tau}) \simeq 3.7$MeV [33]

[64] Electron position annihilation proceeds in states of equilibrium, since it could be reverse by re-contracting and heating up the universe around the transition temperature. Therefore the total entropy is conserved

[65] Notice that protons and neutrons are very non-relativistic (GeV$\gg$MeV) and so can be completely neglected. Electrons and positrons conversely are quasi relativistic

redshifts such that Ta is constant, so they maintain the same temperature as photons until e^+ - e^- annihilation at around 0.5 MeV. After the annihilation, the total entropy is given by[66]

$$\frac{45}{2\pi^2} a_2^3 s_2 = 2(a_2 T_\gamma)^3 + \frac{7}{8} 2 \times 3(a_2 T_\nu)^3 \,, \tag{6.60}$$

where now we accounted for the fact that the neutrinos are not in equilibrium with the photons and so could and indeed have a different temperature T_ν. We can use the conservation of entropy $a_1^3 s_1 = a_2^3 s_2$ to write

$$(a_1 T_1)^3 \frac{43}{4} = 2(a_2 T_\gamma)^3 + \frac{21}{4}(a_2 T_\nu)^3 \,. \tag{6.61}$$

Before the neutrinos decouple, the number of degrees of freedom does not change and so we can use that $T_\nu \propto 1/a$. It turns out that this relation also holds after neutrino decoupling, then the neutrinos are relativistic and out of equilibrium. Hence we can use $a_1 T_1 = a_2 T_\nu$ in (6.61). As a result, a_2 cancels out leaving the following relation between T_ν and T_γ

$$T_\nu = T_\gamma \left(\frac{4}{11}\right)^{1/3} \qquad \Rightarrow \qquad T_\nu(a_0) \simeq 1.96\,\mathrm{K} \simeq 1.7 \times 10^{-4}\,\mathrm{eV} \,. \tag{6.62}$$

So neutrinos are a bit colder than photons at any time after e^+ - e^- annihilation. Notice that this does not depend on whether neutrinos are Dirac or Majorana.

To compute the neutrino energy fraction today Ω_ν, we have to account for their mass. The precise calculation can only be done numerically, but there are two interesting analytical limits. First let us assume the neutrinos are massless. The integral of the Fermi-Dirac distributions is smaller than that over the Bose-Einstein distribution by a factor of 7/8 so we find

$$\rho_\nu = \rho_\gamma \, 3 \, \frac{7}{8} \left(\frac{4}{11}\right)^{4/3} \,, \quad h^2 \Omega_\nu = 1.7 \times 10^{-5} \quad (m_\nu = 0) \,. \tag{6.63}$$

At early times, when neutrinos are still relativistic ($z \gg 500$), the total radiation energy density ρ_r is given by

$$\rho_r = \rho_\gamma \left[1 + N_{\text{eff}} \frac{7}{8} \left(\frac{4}{11}\right)^{4/3} \right] \,, \tag{6.64}$$

where N_{eff} quantifies the number of relativistic species in the universe besides the photons. In the standard model, $N_{\text{eff}} = 3.04$ for the three neutrino species. The slight deviation from 3 comes from the fact that neutrinos still have some small interaction with the SM at e^+ - e^- annihilation and so receive a tiny bit of heating as well. With this result at hand we can compute the the number of relativistic degrees of freedom $g_\star$ below the electron threshold, i.e. after electron positron annihilation,

$$g_* = 2 + \frac{7}{8} (3.04 \times 2) \left(\frac{T_\nu}{T_\gamma}\right)^4 \simeq 3.38 \,. \tag{6.65}$$

[66]The number density of surviving electrons is small and the same as for protons, leading to a charge-neutral universe, $n_e = n_p \sim 10^{-9} n_\gamma$.

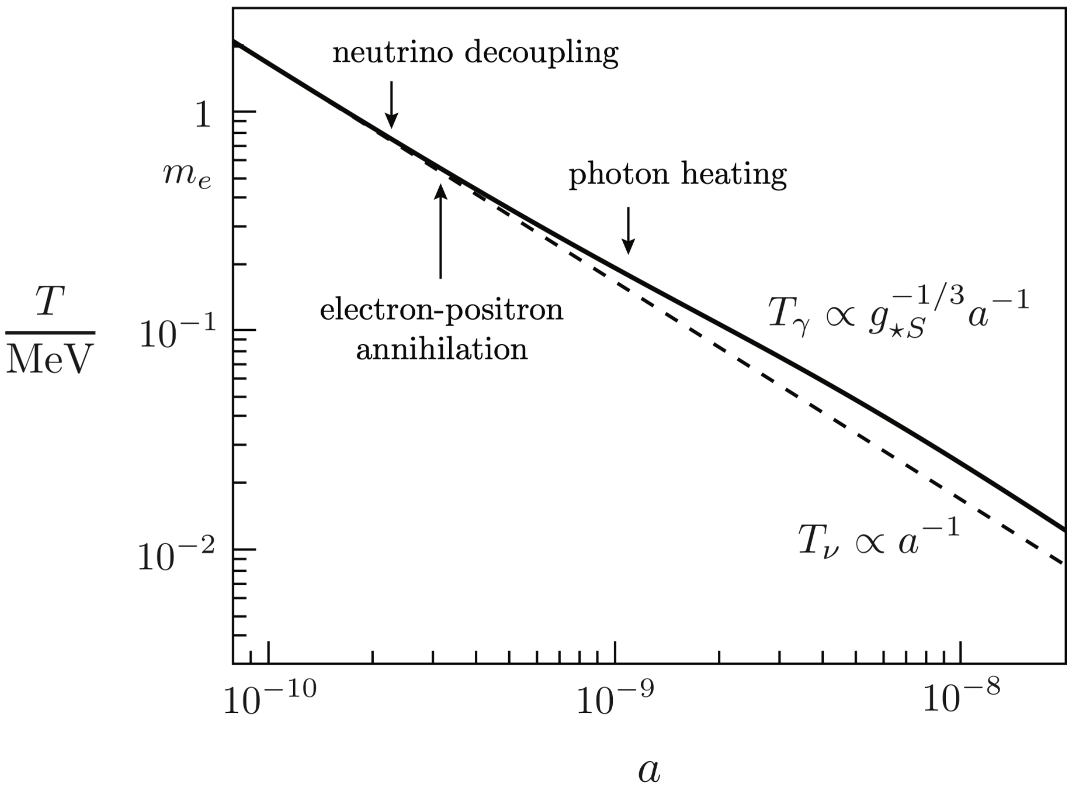


Figure 15: The evolution of the temperature of neutrinos and photons as function of the scale factor a, a proxy for time, around the time of neutrino decoupling and electron-positron annihilation. Credit D. Baumann.

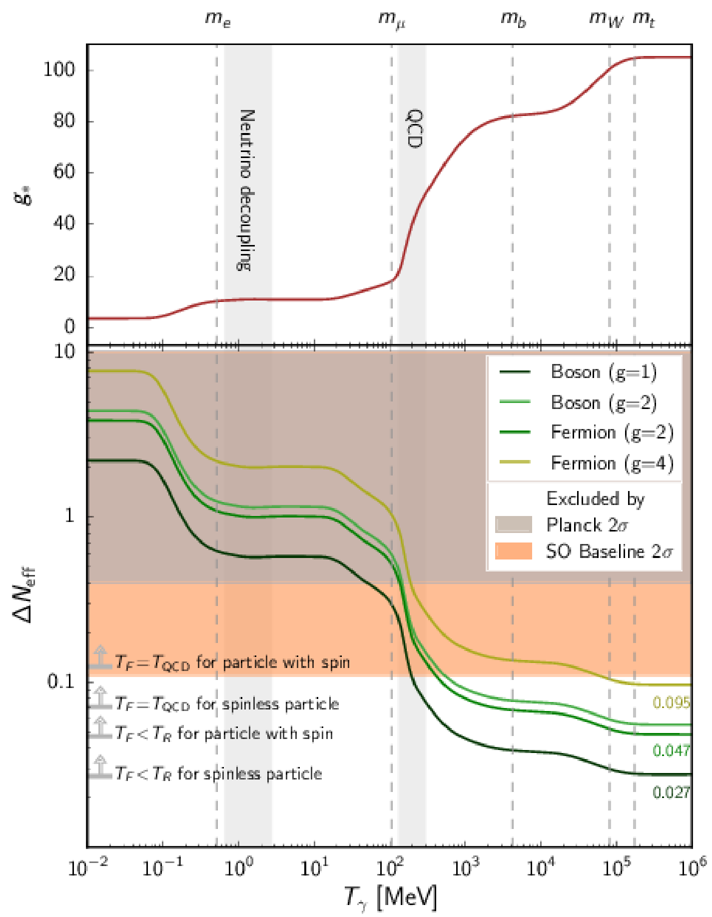


Figure 16: *Top:* The effective number of relativistic species $g_\star$ for the standard model at different temperatures T_γ, which falls from $g_\star = 106.75$ above the top quark threshold, around $T \sim 200$GeV, to $g_* = 3.38$ below the electron threshold, i.e. electron-positron annihilation around $T \sim 0.1$MeV. *Bottom:* $\Delta N_{\rm eff} \equiv N_{\rm eff} - 3.04$ for various types of particles as function of the temperature T_γ at which they decouple. Figure taken from [3].

where we accounted for the two helicities of the photon and of the 3.04 species of out-of-equilibrium neutrinos at temperature T_ν. This is the asymptotic value visible on the right-hand side of Fig. 14.

In analyzing the data, one can treat $N_{\rm eff}$ as a free parameter to test for deviations from the standard model. Currently CMB data gives the constraint $N_{\rm eff} = 3.04 \pm 0.18$ [1], implying a detection of a Cosmic neutrino Background CνB. Sensitivity to $N_{\rm eff}$ is expected to improve by a factor of three in the next ten years with the Simons Observatory (SO) and even further with CMB Stage 4 (S4) [3]. This could detect or exclude any particle that has *ever* been in thermal equilibrium with the Standard Model (see Fig. 16).

7 Thermal history: out of equilibrium

In the previous section we studied the thermal history of the Universe under the assumption of local thermodynamic equilibrium. However, many of the most interesting cosmological processes occur when interactions become inefficient and species fall out of equilibrium. The appropriate framework to describe these transitions is *kinetic theory*, in which the evolution of the phase-space density $f(\mathbf{x}, \mathbf{q}, t)$ is governed by the Boltzmann equation,

$$\frac{df}{dt} = \text{collisions}, \tag{7.1}$$

where the total time derivative must be taken after evaluating f on the solution $\mathbf{x}(t)$ and $\mathbf{q}(t)$ of the equation of motion of the relevant particles. The left hand side is simply the change of the phase space density along a solution of the equations of motion and would vanish by Liouville's theorem in the absence of collisions. The right-hand side accounts for the fact that a collision can change the momentum of a particle and hence the value of f. This most general form of the Boltzmann equation is a partial differential equation in time, space and momentum, encoding how expansion and microscopic interactions compete to shape the distribution of particles. In practice this is very hard to solve. Fortunately, we are often interested not in the full momentum dependence of f but in integrated quantities such as the number density n, the energy density ρ or the momentum flux at position $\mathbf{x}$ and time t, irrespectively of the momentum $\mathbf{x}$. By integrating the Boltzmann equation over d^3q, a process known as "taking moments", one obtains an infinite series of partial differential equation is only $\mathbf{x}$ and time t. Under the assumption of homogeneity and isotropy, these reduce to ordinary differential equation in time since there is no $\mathbf{x}$ dependence. The problem remain non-trivial especially where there are different species of particles that can collide with and transform into each other.

In this section we will first integrate the Boltzmann equation over momenta to derive an evolution equation for the number density of different species, which will allow us to describe particle freeze-out and the associated departure from equilibrium. Then we will apply this formalism to three distinct phenomena: big bang nucleosynthesis, recombination and dark matter decoupling. In the second part of these notes, when considering the inhomogeneous Universe, we will take higher moments of the Boltzmann equation and obtain the continuity and Euler equations that govern the motion of cosmological fluids in space and time.

7.1 The Boltzmann equation

Our goal is to derive an equation to describe the evolution out of chemical equilibrium at various stages in the history of the universe. We will assume isotropy and homogeneity throughout. We will consider exclusively two-to-two scattering, and use the notation $1 + 2 \leftrightarrow 3 + 4$ for the reaction of states or particles 1 through 4. More general processes such as decays or three-body interactions can be described with a similar formalism. Whatever reaction we consider can take place in both directions. The variables we want to describe are the densities of the four species as function of time in a flat FLRW universe. The *integrated Boltzmann equation* *for annihilation* is given by[67]

$$a^{-3}\frac{d\left(a^3 n_1\right)}{dt} = \int \prod_{i=1,4} \frac{d^3 p_i}{(2\pi)^3 2E_i} \delta_D^3\left(\sum_{i=1,4} \vec{p}_i\right) \delta_D\left(\sum_{i=1,4} E_i\right) |M|^2$$
$$\left[f_3 f_4 \left(f_1 \pm 1\right)\left(f_2 \pm 1\right) - f_1 f_2 \left(f_3 \pm 1\right)\left(f_4 \pm 1\right)\right] , \quad (7.2)$$

and similarly for particles 2 through 4. Several comments are in order:

- In the absence of any interaction, the right-hand side vanishes and $a^3 n_1$ is constant, as expected in an expanding universe if n_1 is a conserved charge density. Using the notation of (1.40) this can also be written as

$$0 = (u^\mu n)_{;\mu} = a^{-3}\frac{d\left(a^3 n_1\right)}{dt} . \quad (7.3)$$

- $f_i(\mathbf{x}, \vec{p}, t)$ is the phase space density function. In this section we will assume it does not depend on space, but we will amend this when studying the inhomogeneous universe.
- The factors $(f_i \pm 1)$ come about because of the quantum statistic and are called respective Bose enhancement and Pauli blocking. Bose enhancement means that it is easier to produce a boson in a state that is already occupied by a large number of particles. Pauli blocking means that for fermions the density of state cannot be larger than one.
- The term $f_3 f_4 \left(f_1 \pm 1\right)\left(f_2 \pm 1\right)$ describes particles 3 and 4 combining to create particles 1 and 2, while the second term describes the opposite process.
- According to quantum mechanics the complex *probability amplitude* M is used to compute the real probability $|M|^2$ of the process to happen. This is the only place where the dynamics of the theory under consideration appears. M is proportional to the coupling constant responsible for the interaction.
- The Dirac deltas ensure energy and momentum conservation in each interaction.
- The integrals over all four momenta sum over all possible ways that the interaction can proceed.
- This expression is *not* invariant under time reversal T. This is related to the surprising fact that even when microscopic physics is invariant under T, macroscopic process are observed to have an "arrow of time". In deriving this equation from the BBGKY hierarchy[68] we have neglected the 2-, 3-, . . . n-particle densities and therefore we have lost the correlation among particles that is generated by interactions. This loss of information breaks T, even if the underlying interactions were T-symmetric, and lead to the possible increase of entropy and selects an arrow of time.
- This is a coupled system of non-linear, ordinary, integro-differential equations, a.k.a. it is pretty hard to solve.

[67] CFU: *Why the factor of* $1/E_i$*? To make the measure Lorentz invariant. It can be alternatively written as* $\int d^3 p_i dE_i\, \delta_D(E_i^2 - p_i^2 - m_i^2)$*, with* m_i *the mass of the* i*-th particle*

[68] See chapter 68 of [48] for a clear discussion.

To make some progress we will make two simplifying assumptions. The first is that there is *kinetic equilibrium*. This means that there are efficient interactions that distribute energy and momentum within a single species very quickly. It implies that we can use the Bose-Einstein or Fermi-Dirac equilibrium distribution for the distribution functions f_a. Notice that this is *not* the same as *chemical equilibrium*, in which there are efficient interactions to change particles from one species to another. For example, in chemical equilibrium $\mu_1 + \mu_2 = \mu_3 + \mu_4$, but we will not assume this in the following. Intuitively this means that we consider a situation in which particles can change their energy and momentum but not necessarily their type. The second assumption, which is valid for the three cases of interest to us, is that $T \ll E - \mu$ and therefore we can drop the ± 1 in the Fermi-Dirac and Bose-Einstein distributions and simply use the classical Boltzmann equilibrium distribution f_B,

$$f_{BE,FD} = \frac{1}{e^{(E-\mu)/T} \mp 1} \simeq e^{-(E-\mu)/T} = f_B \,. \tag{7.4}$$

This also implies we can approximate the Bose-enhancement and Pauli-blocking factors by $f_i \pm 1 \simeq \pm 1$ since $e^{-(E-\mu)/T} \ll 1$.
Recall that the species' out-of-equilibrium densities n_i were given in the previous section by

$$n \equiv g \int \frac{d^3p}{(2\pi)^3}\, e^{-(E-\mu)/T} = g\, e^{\mu/T} \begin{cases} \left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T} & T \ll m \\ \frac{T^3}{\pi^2} & T \gg m \end{cases} \,, \tag{7.5}$$

and the species' equilibrium densities $n_i^{(0)}$ where

$$n^{(0)} \equiv n|_{\mu=0} = n\, e^{-\mu/T} \,. \tag{7.6}$$

Assuming the chemical potential is momentum-independent, we can then write

$$\left[f_3 f_4 \left(f_1 \pm 1\right)\left(f_2 \pm 1\right) - f_1 f_2 \left(f_3 \pm 1\right)\left(f_4 \pm 1\right)\right] \simeq e^{-(E_1+E_2)/T} \left[\frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} - \frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}}\right] , \tag{7.7}$$

where we used the conservation of energy $E_1 + E_2 = E_3 + E_4$. Upon defining the *thermally averaged cross section* as

$$\langle \sigma v \rangle \equiv \frac{1}{n_1^{(0)} n_2^{(0)}} \int \prod_{i=1,4} \frac{d^3 p_i}{(2\pi)^3 2E_i} \delta_D^3 \left(\sum_{i=1,4} \vec{p}_i\right) \delta_D \left(\sum_{i=1,4} E_i\right) |M|^2 e^{-(E_1+E_2)/T} \,, \tag{7.8}$$

we can finally write our final form of the integrated Boltzmann equation

$$a^{-3} \frac{d\left(a^3 n_1\right)}{dt} = \langle \sigma v \rangle n_1^{(0)} n_2^{(0)} \left[\frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} - \frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}}\right] . \tag{7.9}$$

Upon including the other Boltzmann equations obtained by permuting the species indices, one obtains a coupled set of ordinary differential equations. Since $[n] = L^{-3}$ and $[\langle \sigma v \rangle] = L^3 t^{-1}$, we can think of $n_2^{(0)} \langle \sigma v \rangle$ as a reaction rate Γ, with units of inverse time $[\Gamma] = t^{-1}$. Here, Γ is rate at which species 1 is created or destroyed due to the reaction $1 + 2 \leftrightarrow 3 + 4$. Expanding out the derivative on the left-hand side we can write the Boltzmann equation in the suggestive form

$$\dot{n}_1 = -3Hn_1 + \Gamma n_1^{(0)} n_2^{(0)} \left[\frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} - \frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}} \right] . \tag{7.10}$$

Depending the rate of interaction Γ and rate of expansion of the universe H, there are two regimes:

- *Equilibrium*, $\Gamma \gg H$: the reaction is very efficient and determines the relative densities of species. Generic initial values for n_i are very quickly driven to the chemical equilibrium

$$\frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} = \frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}} \quad \Leftrightarrow \quad \mu_1 + \mu_2 = \mu_3 + \mu_4 \,, \tag{7.11}$$

  which ensures a large cancellation of the individual terms on the right-hand side. In the cosmology context, the above equation is sometimes called the Saha equation. Notice that, even if $\mu_i \neq 0$, the ratio of abundances is the same as it would be if $\mu_i = 0$ for every i, namely

$$\frac{n_3 n_4}{n_1 n_2} = \frac{n_3^{(0)} n_4^{(0)}}{n_1^{(0)} n_2^{(0)}} . \tag{7.12}$$

- *Freeze-out*, $\Gamma \ll H$: the reaction is too slow to keep up with the relatively faster expansion of the universe. One can neglect the right-hand side of Eq. (7.9) and find $n_i(t) \simeq n_i(a_*)(a_*/a)^{-3}$, where $*$ refers to the last moment at which $\Gamma \simeq H$. In particular, after an interaction goes out of equilibrium the ratio of all species involved becomes constant, assuming there aren't other processes that affect them. This is called "freeze-out".

In the following we will use the above formalism to study three distinct processes: big bang nucleosynthesis, recombination and dark matter decoupling.

7.2 Big bang nucleosynthesis

After the QCD phase transitions all quarks confined into the two long-lived baryons: protons and neutrons. The baryons in the universe today are observed to be 75% Hydrogen (^{1}H) and 25% Helium (^{4}He) with only traces of the other isotopes and of the heavier elements including us, see Fig. 18. Given that the typical binding energy of these nuclei is a few MeV per nucleon, at temperatures much above a few MeV, but still well below the QCD phase transition, all the baryons in the universe were free protons and neutrons. At these temperatures, any nucleus that formed would be instantaneously destroyed by some MeV photon in the thermal bath. So it is natural to ask how the observed abundance of elements arose as the universe expanded and cooled below this temperature[69]. This is the goal of *Big Bang Nucleosynthesis (BBN).*

To study BBN analytically, we will decompose the problem in two separated steps:

[69] It is straightforward to check (see P.7.1) that only a small fraction of He could have been synthesized in stars at later times.

1. Calculation of the free neutron abundance, $T > 0.1$ MeV. To obtain a reasonable estimate we will need to solve the Boltzmann equation out of equilibrium.

2. Formation of Deuterium, Helium and heavier atoms, $T < 0.1$ MeV. Here it will suffice to assume thermodynamical equilibrium and estimate the temperature T_{nuc} when all neutrons swiftly combine into Deuterium first and Helium-4 immediately afterwards.

This is a well justified separation for an estimate because the creation of nuclei is heavily suppressed above 0.1 MeV.

Free neutron abundance Above $T = 0.1$ MeV almost all neutrons and protons are free, i.e. they have not combined to form atomic nuclei. Free neutrons and protons can turn into each other via the weak interactions. At MeV energies the effective Fermi theory contains the following two-body processes[70]

$$n + \nu_e \leftrightarrow p^+ + e^- , \quad n + \bar{e}^+ \leftrightarrow p^+ + \bar{\nu}_e \,. \tag{7.13}$$

We neglect for the moment neutron decay via the process $n \to p^+ + e^- + \bar{\nu}_e$, to which we will come back later. If free protons and neutrons remained in chemical equilibrium, i.e. $\mu_p = \mu_n = 0$, their ratio would be simply set by the equilibrium expression in Eq. (7.5),

$$\frac{n_n}{n_p} = \frac{n_n^0}{n_p^0} = \frac{\left(\frac{m_n T}{2\pi}\right)^{3/2} e^{-m_n/T}}{\left(\frac{m_p T}{2\pi}\right)^{3/2} e^{-m_p/T}} \simeq e^{(m_p - m_n)/T} \equiv e^{-Q/T} \,, \tag{7.14}$$

where we used that $m_n/m_p \simeq 1.001$ and we introduced the mass difference is

$$Q \equiv m_n - m_p \simeq 1.3 \text{ MeV} \,. \tag{7.15}$$

The equilibrium solution in (7.14) tells us that almost all neutrons would have turned into protons shortly after $T = 1.3$ MeV. To quantify this we define the fraction X_n of baryons in neutrons by

$$X_n \equiv \frac{n_n}{n_n + n_p} . \tag{7.16}$$

In chemical equilibrium we would have $X_n = 1/2$ at high temperatures and early times, where $e^{-Q/T} \sim 1$ and $X_n \sim e^{-Q/T}$ at low temperatures and late times, where $e^{-Q/T} \ll 1$. Luckily for us, whose existence relies on atoms heavier than hydrogen, the weak interactions responsible for converting neutrons into protons go out of equilibrium and some relic neutrons survive. A precise derivation requires working outside of chemical equilibrium by solving the differential Boltzmann equation Eq. (7.9). However, the general features can already be appreciated with a simple equilibrium calculation. EP: *To be written*

[70] Three and higher n-body processes are suppressed when the number densities n are low with respect to the typical interaction volume.

We start by assuming that the leptons are in complete chemical equilibrium and so $n_l = n_l^{(0)}$. This means that the Boltzmann equation, Eq. (7.9), becomes

$$a^{-3}\frac{d\left(a^3 n_n\right)}{dt} = n_l^{(0)}\langle\sigma v\rangle\left[\frac{n_p n_n^{(0)}}{n_p^{(0)}} - n_n\right]. \tag{7.17}$$

Next, we rewrite the neutron density n_n in terms of the dimensionless fraction X_n. The clever way to do this is is to notice that for all the weak processes in (7.13), the total number of baryons is conserved, so $(n_p + n_n)a^3$ is constant. Writing

$$X_n = \frac{n_n a^3}{(n_n + n_p)a^3}, \tag{7.18}$$

we can massage the left-hand side of (7.17) into

$$\frac{1}{a^3}\frac{d(a^3 n_n)}{dt} = \frac{1}{a^3}\frac{d((n_n+n_p)a^3 X_n)}{dt} = (n_n + n_p)\dot{X}_n = n_n\frac{\dot{X}_n}{X_n}. \tag{7.19}$$

Next we introduce the neutron-proton conversion rate $\lambda_{np}(T) = n_l^{(0)}\langle\sigma v\rangle$, which is temperature dependent. The calculation in quantum field theory outlined in [] gives

$$\lambda_{np} = \frac{255}{t_{\rm life}}\left[12\left(\frac{T}{Q}\right)^5 + 6\left(\frac{T}{Q}\right)^4 + \left(\frac{T}{Q}\right)^3\right], \tag{7.20}$$

with $t_{\rm life} = 887\text{sec} \sim 15$ minutes. Using $n_n^{(0)}/n_p^{(0)} = e^{-Q/T}$, the right-hand side of (7.17) can be written as

$$\lambda_{np} n_n\left[\frac{n_p n_n^{(0)}}{n_n n_p^{(0)}} - 1\right] = \lambda_{np} n_n\left[\left(\frac{1}{X_n} - 1\right)e^{-Q/T} - 1\right]. \tag{7.21}$$

Equating left- to righ-hand side we find

$$\dot{X}_n = \lambda_{np}\left[(1 - X_n)\,e^{-Q/T} - X_n\right]. \tag{7.22}$$

It will be informative to use temperature as a proxy for time by writing

$$\frac{d}{dt} = \dot{T}\frac{d}{dT} = -HT\frac{d}{dT}. \tag{7.23}$$

To solve this we must work out the T dependence of H. This is given by Friedmann equation

$$H = \sqrt{\frac{\rho}{3M_{\rm Pl}^2}} = \frac{\pi}{3}\sqrt{\frac{g_\star}{10}}\frac{T^2}{M_{\rm Pl}}, \tag{7.24}$$

where we choose $g_* \simeq 10.75$, consisting of photons, three families of left-handed neutrinos and their anti-particle, and the left and right-handed electrons and their anti-particles the positrons. The final equation

$$T\frac{dX_n}{dT} = -\frac{\lambda_{np}}{H}\left[(1 - X_n)\,e^{-Q/T} - X_n\right] \tag{7.25}$$

can be solved numerically and the solution is shown in Fig. 17. As we will see shortly we will need the value of X_n at $T = 0.1$ MeV. Using the numerical solution to the above equation this is found to be $X_n(0.1\text{ MeV}) = 0.11$.

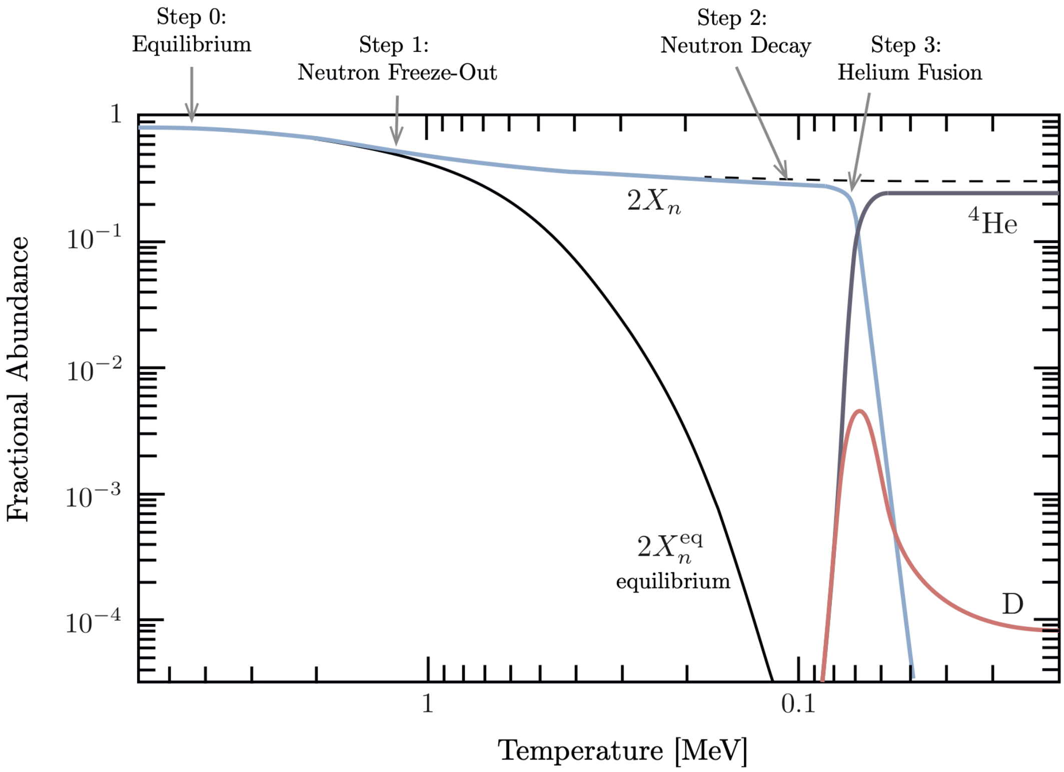


Figure 17: Twice the fractional abundance of neutrons over baryons $2X_n$ as function of temperature. Around $T \simeq 1$MeV neutrons go out of equilibrium and freeze out shortly thereafter to $X_n \simeq 0.15$. Almost all free neutrons at $T = 0.1$ MeV are quickly converted into Helium-4. The remaining neutrons either decay or combine to produce a small but measurable amount of Deuterium and Helium-4. Figure from D. Baumann.

Light element formation A good approximation to light element formation is that it happens instantaneously at some temperature T_{nuc} that can be calculated in equilibrium. By the end of BBN, almost all of the neutrons are transformed into Helium-4, with only traces of Deuterium, Helium-3 and Lithium. Since it is unlikely for two free protons and two free neutrons to find each other and combine into Helium-4 in one go, the reaction passes through the temporary creation of Deuterium, which then in turn fuses into Helium-4. To quantify the creation of Deuterium, we note that its equilibrium abundance is determined by the nuclear process

$$n + p \leftrightarrow D + \gamma \,. \tag{7.26}$$

Assuming that this process takes place in equilibrium, we can simply demand the vanishing of the right-hand side of the Boltzmann equation (7.9) adapted to the above process, namely

$$\left[\frac{n_D n_\gamma}{n_D^{(0)} n_\gamma^{(0)}} - \frac{n_n n_p}{n_n^{(0)} n_p^{(0)}} \right] \simeq 0 \,. \tag{7.27}$$

Since photons have negligible chemical potential $n_\gamma = n_\gamma^{(0)}$, and so (see P.7.6)

$$\frac{n_D}{n_n n_p} = \frac{n_D^{(0)}}{n_n^{(0)} n_p^{(0)}} \,. \tag{7.28}$$

Using again (7.5) or (6.26), this reduces to

$$\frac{n_D}{n_n n_p} = \frac{3}{4}\left(\frac{2\pi m_D}{m_n m_p T}\right)^{3/2} e^{(m_n+m_p-m_D)/T} \simeq \frac{3}{4}\left(\frac{4\pi}{m_p T}\right)^{3/2} e^{B_D/T} \,, \tag{7.29}$$

where we introduced the binding energy of Deuterium $B_D = m_n + m_p - m_D \simeq 2.2$ MeV and approximated $m_D \sim 2m_p \sim 2m_n$ in the fraction. We now drop order one factors, approximate $n_n \sim n_p \sim n_b$ and introduce the *baryon-to-photon* number ratio η_b, which is measured to be

$$\eta_b \equiv \frac{n_b}{n_\gamma} = \frac{\rho_b}{m_p} \frac{\pi^2}{2\zeta(3) T_{\rm CMB}^3} \simeq 5 \times 10^{-10} \,. \tag{7.30}$$

Finally, (7.29) becomes

$$\frac{n_D}{n_b} \simeq \eta_b \left(\frac{T}{m_p}\right)^{3/2} e^{B_D/T} \,, \tag{7.31}$$

The physics behind this equation is that the process $D + \gamma \leftrightarrow p + n$ happens in both ways as long as there are enough photons with energy larger than 2.2 MeV, which are able to break up Deuterium. Naively one would expect this to stop being true at a temperature around 2.2 MeV. This rough estimate neglects the fact that for every baryon there are a billion photons, Eq. (7.30). Even when $T < 2.2$ MeV, there are still enough photons in the high-energy tails of the phase space distribution to destroy Deuterium. As a result, Deuterium remains in equilibrium well past the temperature of 2.2 MeV.

As more Deuterium is produced, the reservoir of free neutrons is depleted. This proceeds until the density of free neutrons is so low that chemical equilibrium cannot be maintained. A rough estimate of the temperature T_{nuc} at which chemical equilibrium is lost is obtained by noticing that, since baryon number is conserved, the exact solution must obey $n_D \leq n_b$. But chemical equilibrium gives an $n_D(T)$ that grows well past n_b, and so $n_D(T_{nuc}) = n_b$ is a good estimate of when that approximation breaks down. Equating (7.31) to one and solving for $T = T_{nuc}$ one finds

$$T_{nuc} \gtrsim \frac{2.2}{\log(5 \times 10^{-10})} \text{MeV} \sim \frac{2.2}{20} \text{MeV} \simeq 0.1 \text{MeV} \,. \tag{7.32}$$

This tells us that around $T = 0.1$ MeV all remaining neutrons have combined with protons to form deuterium. Next we notice that, since the binding energy for Helium is higher than that of Deuterium, as soon as Deuterium is present it immediately fuses into Helium-4. As a result, vary soon after T_{nuc}, the Helium abundance grows larger than the deuterium abundance. It is therefore a good approximation to assume all free neutrons have fused into ^{4}He by $T = 0.1$ MeV. Since two neutrons are needed for one Helium nucleus, our prediction for the Helium mass fraction is

$$X_4 \equiv \frac{4 n_{^4He}}{n_b} \approx 2 X_n(T_{nuc}) = 0.22. \tag{7.33}$$

This quantity is also often denotes as Y_p, where the letter Y refers to Helium and the label p stands for primordial. As it turns out, this is a remarkably good approximation to the actual abundances observed. The prediction that roughly one quarter of the baryonic mass of the Universe should be in the form of Helium-4 was one of the first quantitative successes of the hot Big Bang model. When this calculation was first performed in the late 1940s by Alpher, Bethe and Gamow, no other cosmological model of the time could account for the observed light-element abundances, while stellar processes were clearly insufficient to produce such large quantities of Helium. The agreement between the predicted and measured primordial fractions of ^{4}He, together with the later confirmation of the cosmic microwave background, established the Big Bang theory as the standard description of the early Universe and made BBN one of its earliest and most robust observational pillars.
Beyond its historical role, BBN remains an essential probe of the baryon content of the Universe. The freeze-out abundances of light nuclei depend sensitively on the baryon-to-photon ratio η_b, since a higher baryon density enhances nuclear reaction rates and shifts the temperature at which the light elements form. In particular, the residual abundance of Deuterium is an excellent "baryometer": its measured value in high-redshift absorption systems tightly constrains η_b and therefore the present-day baryon density $\Omega_b h^2$. The most precise analyses give

$$\Omega_b h^2 = 0.0205 \pm 0.0018, \tag{7.34}$$

in excellent agreement with, and completely independent from, the determination obtained from the CMB. This concordance between two very different physical epochs—minutes and hundreds of thousands of years after the Big Bang—is one of the strongest internal consistency checks of the standard cosmological model.

7.3 Recombination*

Process:

$$e^- + p^+ \leftrightarrow H^0 + \gamma. \tag{7.35}$$

Happens around ~ 1 eV. Again compare with hydrogen binding energy of 13.6 eV. We are going to track the electron to hydrogen abundance (so the effect of the expansion of the universe drops out)

$$X_e \equiv \frac{n_e}{n_e + n_H} = \frac{n_p}{n_p + n_H}, \tag{7.36}$$

where the latter equality is ensured by the neutrality of the universe. Note that this is in principle not the quantity that measures the kinetic equilibrium of the photons: the photons are chemically decoupled from the fluid since 1 MeV, but remain in kinetic equilibrium until decoupling of photons from matter, which we investigate later. Due to the imbalance between photons and baryons, the decoupling of matter from photons is yet another question. Our discussion of recombination will be very similar to the neutron-to-proton ratio story. Once again, one can start from equilibrium considerations to find the temperature at which the ratio starts to change significantly. The equilibrium condition (Saha equation) is again

$$\frac{n_e n_p}{n_H} = \frac{n_e^{(0)} n_p^{(0)}}{n_H^{(0)}}, \tag{7.37}$$

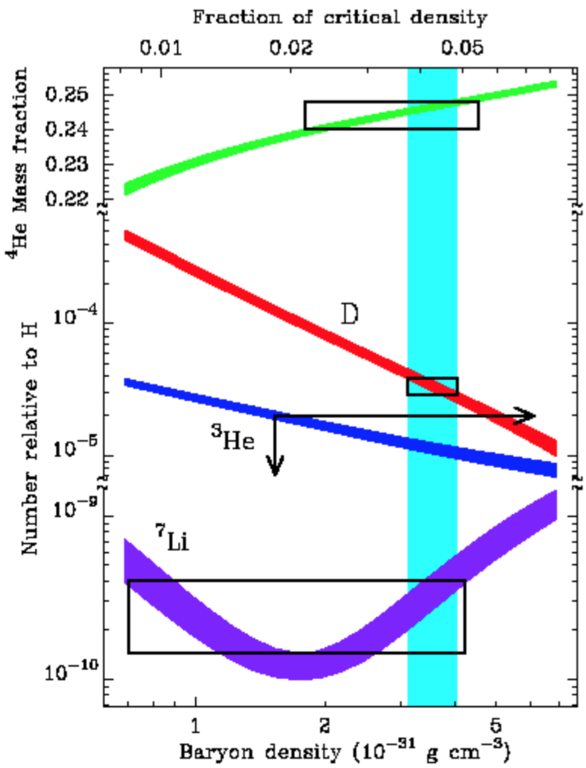


Figure 18: The abundance of the lightest elements (Deuterium, Helium 3 and 4, Lithium) as function of the baryon density today as predicted by BBN (colored bands). Black boxes represent the observational constraints. For $\Omega_b = 0.04$ all data are compatible with predictions.

which can be written as
The equation governing the electron fraction going out of equilibrium is

$$\frac{dX_e}{dt} = \langle\sigma v\rangle \left\{ (1 - X_e) \left(\frac{m_e T}{2\pi}\right)^{3/2} e^{-\epsilon_0/T} - X_e^2 n_b \right\}. \tag{7.38}$$

Very similar to previous case. Difference is in the fact that electron mass matters and $n_e = n_p$ (explains the square appearing) and we use $n_e + n_H = n_b$. For the cross section we need

$$\langle\sigma v\rangle = \alpha^{(2)} = \frac{10\alpha^2}{m_e^2} \left(\frac{\epsilon_0}{T}\right)^{1/2} \ln\left(\frac{\epsilon_0}{T}\right). \tag{7.39}$$

Draw energy levels of hydrogen.
[71]

Fig. 19 show the result of the numerical integration.

Decoupling Photons remain in kinetic eq. mainly due to Thomson scattering off electrons, $\sigma_T \simeq 0.7 \times 10^{-24}$ cm^{-2}. They go out of equilibrium when this rate becomes comparable to the

[71] CFU: *Why $\alpha^{(2)}$? Answer: 13.6 eV photon is reionizing. Formation through cascade.*

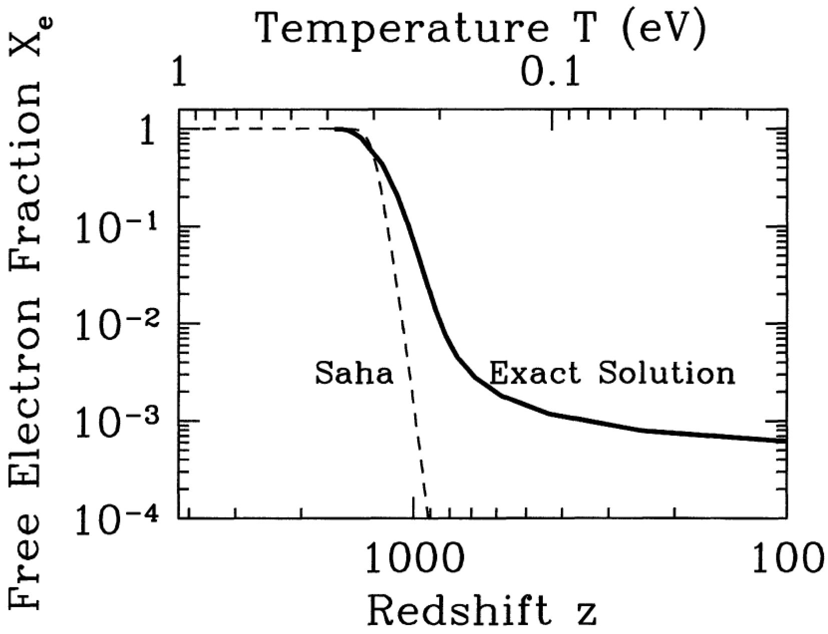


Figure 19:

expansion rate of the universe:

$$n_e \sigma_T = X_e n_b \sigma_T \sim H^{-1}. \tag{7.40}$$

Plugging in the numbers we find [Problem P.7.7]

$$\frac{n_e \sigma_T}{H} = 113 X_e \left(\frac{\Omega_b h^2}{0.02}\right) \left(\frac{0.15}{\Omega_m h^2}\right)^{1/2} \left(\frac{1+z}{1000}\right)^{3/2} \left[1 + \frac{1+z}{3600}\frac{0.15}{\Omega_m h^2}\right]. \tag{7.41}$$

7.4 Dark matter decoupling*

We investigate the WIMP scenario here. There are other scenario's, such as decaying DM, which allow one to search for DM masses in a wide range. The WIMP scenario however predicts GeV masses. Process:

$$X + X \leftrightarrow l + l, \tag{7.42}$$

where X is the heavy DM particle and l is a known light particle that DM weakly interacts with. The light particles are in complete chemical as well as kinetic equilibrium. The equation governing the DM fraction

$$Y \equiv \frac{n_X}{T^3}, \tag{7.43}$$

now becomes

$$\frac{dY}{dt} = T^3 \langle \sigma v \rangle \left\{ Y_{EQ}^2 - Y^2 \right\}, \tag{7.44}$$

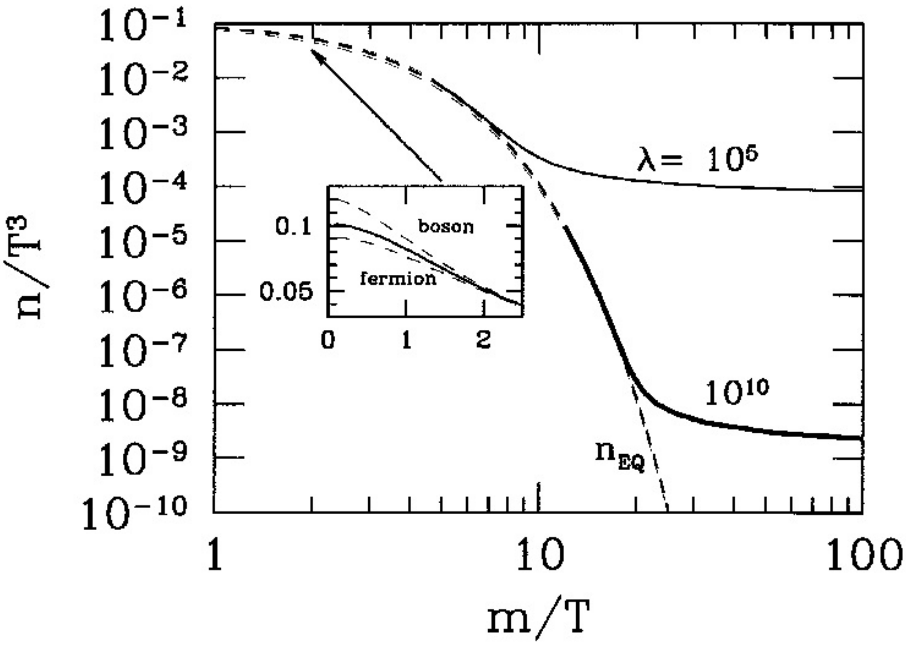


Figure 20:

with $Y_{Eq} \equiv n_X^{(0)}/T^3$. At very high temperatures $T \gg m$, dark matter was relativistic and $Y_{Eq} = 1$. Note that when the process goes out of equilibrium, $Y > Y_{EQ}$, and Y decreases with time, so the sign in this equation is correct. The T^3 comes from the fact that the relevant era here is radiation, during which $a \sim T^{-1}$. Again we would need to know the temperature dependence and size of the cross section when the temperature is of the order of the DM particle mass. (Comment on approximation for and definition of λ?) Let us use $x \equiv m/T$ as time, then we find the master equation

$$\frac{dY}{dx} = -\frac{\lambda}{x^2}\left[Y^2 - Y_{Eq}^2\right] , \tag{7.45}$$

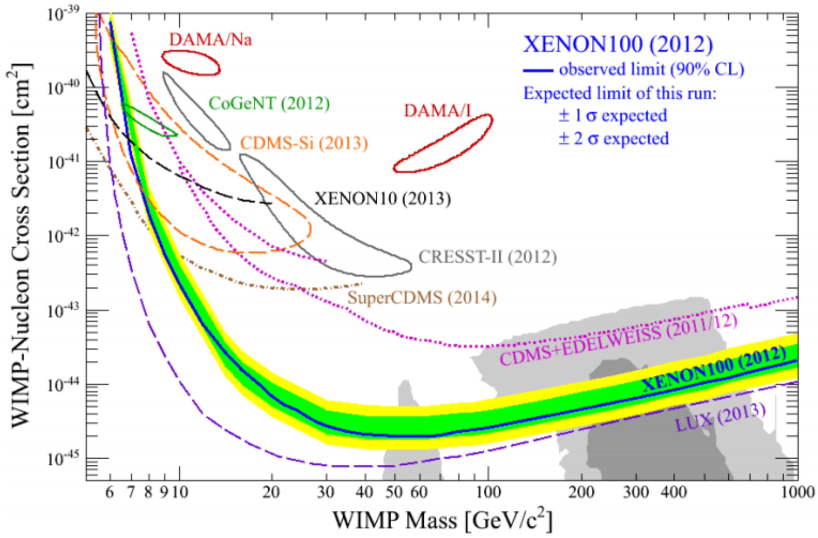
with $\lambda \equiv m^3 \langle \sigma v \rangle / H(m)$. After looking at the numerical solution in Fig. 20, we can derive a rough estimate of the freeze out abundance

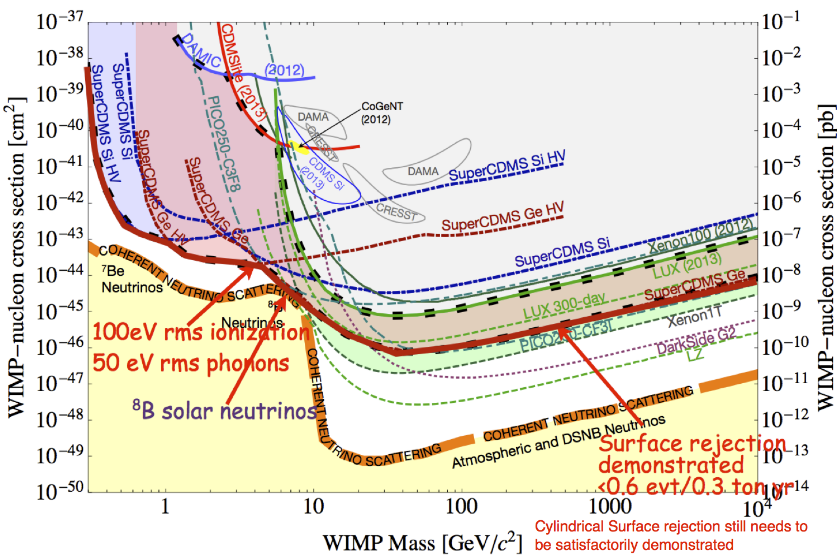
$$\frac{dY}{dx} \sim -\frac{\lambda}{x^2} Y^2 \quad \Rightarrow \quad Y_\infty \sim \frac{x_f}{\lambda} \sim \frac{10}{\lambda} . \tag{7.46}$$

After freeze-out, the energy density falls off as $1/a^3$. However, just like for neutrinos, the photon fluid temperature develops slightly differently due to the decoupling of all massive particles in the range 100 GeV till now. Therefore the DM energy density today is

$$\rho_X = m Y_\infty T_0^3 \left(\frac{a_1 T_1}{a_0 T_0}\right)^3 \approx m Y^\infty T_0^3/30. \tag{7.47}$$

where $(a_1T_1)/(a_0T_0) \sim 1/30$ arises from the fact the number of degrees of freedom around $T \sim GeV$ was about 100, while it is a few today [Problem P.7.4] Then one can make prediction





of Ω_X: insensitive to DM mass, since $Y_\infty \sim x_f \sim 1/m$: energy density does not depend on mass (apart from indirect dependence of $g_\star$ and the final ratio m/T_f). WIMP mass high compared to SM particles. Then we can estimate the relevant cross sections: a few orders of magnitude below estimates from supersymmetry. Show plots.

Problems for lesson 7

P.7.1 Compute the luminosity-to-mass ratio that stars would need in order to synthesize the observed 25% of ^{4}He during the last 14 billion years. Compare the result to the observed luminosity-to-mass (baryonic) ratio observed in the universe, $L/M_b \lesssim 0.05 L_☼/M_☼$, where the label ☼ refer to our sun. [Hint: compute the energy per baryon from the binding energy of He. Divide by the age of the universe and compare with $L_☼/M_☼$]

P.7.2 (From Dodelson Ch 3, Ex 2) Track the density of electron and positron during BBN. Since electromagnetic interactions are very strong during BBN, you can estimate this using $\mu_{e^-} = \mu_{e^+} = \mu_\gamma = 0$. When does the energy density of n_e fall to 1% of that of photons?

P.7.3 (From Dodelson Ch 3, Ex 6) Determine the baryon to photon ratio and show it is approximately given by

$$\eta_b \equiv \frac{n_b}{n_\gamma} = 5 \times 10^{-10} \left(\frac{\Omega_b h^2}{0.02} \right) \tag{7.48}$$

P.7.4 (from Dodelson Ch 3, Ex 11) As long as g_* is constant, the conservation of total entropy $s \propto a^{-3}$ plus the relation $s \propto g_* T^3$ (since entropy is dominated by relativistic species) implies $T \propto 1/a$. Compute aT at $T = 10$ GeV and today to quantify how much our universe deviates from the simple inverse scaling relation, due to the change in g_*.

P.7.5 Exercises 7 (D3) on the baryon loading for recombination (don't forget neutrinos)

P.7.6 *Optional* Estimate the Deuterium abundance assuming chemical equilibrium Eq. (7.32). [Hint: use the chemical equilibrium condition Eq. (7.11) and the binding energy of Deuterium $m_p + m_n - m_D \simeq 2.2$ MeV.]

P.7.7 Derive equation Eq. (7.41)

Part II
The inhomogeneous Universe

8 Cosmological perturbation theory

In this section, we sail off the land of calm and homogenous seas into the perilous and stormy spacetime oceans. More specifically we assume

$$g_{\mu\nu}(x,t) = \bar{g}_{\mu\nu}(t) + h_{\mu\nu}(x,t)\,, \tag{8.1}$$
$$T_{\mu\nu}(x,t) = \bar{T}_{\mu\nu}(t) + \delta T_{\mu\nu}(x,t)\,, \tag{8.2}$$

with $|h_{\mu\nu}| \ll |\bar{g}_{\mu\nu}|$, $|\delta T_{\mu\nu}| \ll |\bar{T}_{\mu\nu}|$ and barred quantities representing the homogenous and isotropic exact *background* solutions we discussed in the previous sections. In particular, $\bar{g}_{\mu\nu}$ is the flat FLRW metric in (1.15), $\bar{T}_{\mu\nu}$ was given in Eq. (1.27) and $\bar{u}^\mu = \{1,0,0,0\}$. We work perturbatively in the small perturbations $|h_{\mu\nu}|$ and $|\delta T_{\mu\nu}|$. Throughout we will assume a spatially-flat Universe, $K = 0$.

When working with perturbations, one has to decide about the meaning of upper, covariant indices, and lower, contravariant indices. We use the convention that $\delta T^{\cdots}_{\cdots}$ represents the perturbation of $T^{\cdots}_{\cdots}$, as opposed to the perturbations of say $T_{\cdots\cdots}$ raised by the background metric. For example, we define $h^{\mu\nu}$ by

$$h^{\mu\nu} \equiv g^{\mu\nu} - \bar{g}^{\mu\nu}\,, \tag{8.3}$$

where $\bar{g}^{\mu\nu}$ is the inverse of $\bar{g}_{\mu\nu}$ so that $\bar{g}^{\mu\nu}\bar{g}_{\nu\sigma} = \delta^\mu_\sigma$. Given the above definition, $h^{\mu\nu}$ is related to $h_{\mu\nu}$ as follows:

$$(\bar{g}^{\mu\nu} + h^{\mu\nu})(\bar{g}_{\nu\sigma} + h_{\nu\sigma}) = \delta^\mu_\sigma \quad \Rightarrow \quad h^{\mu\nu}\bar{g}_{\nu\sigma} = -\bar{g}^{\mu\nu}h_{\nu\sigma} \quad \Rightarrow \quad h^{\mu\rho} = -\bar{g}^{\rho\sigma}\bar{g}^{\mu\nu}h_{\nu\sigma}\,. \tag{8.4}$$

Very explicitly, for the flat FLRW metric $\bar{g}_{\mu\nu} = \mathrm{Diag}(-1, a^2, a^2, a^2)$, we find

$$h^{00} = -h_{00}\,, \qquad h^{0i} = +\frac{1}{a^2}h_{0i}\,, \qquad h^{ij} = -\frac{1}{a^4}h_{ij}\,. \tag{8.5}$$

Note in particular that $h^{\mu\nu}$ is *not* obtained by raising the indices of $h_{\mu\nu}$ with the background metric.

8.1 Linearised equations of motion

In these notes we mostly focus on the leading non-trivial order, namely linear order in $h_{\mu\nu}$ and $\delta T_{\mu\nu}$. We want to expand all equations of motions to *linear order* in perturbations. We start from the two (dependent) set of equations[72]

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -\frac{1}{M_{\rm Pl}^2}T_{\mu\nu}\,, \quad T^{\mu\nu}{}_{;\nu} = 0\,. \tag{8.6}$$

For our calculation we will employ a slightly different form of these equations. For this, first take the trace Einstein equations with $g^{\mu\nu}$ to find $R = T/M_{\rm Pl}^2$ where $T \equiv T^\lambda{}_\lambda$. Then, the term $1/2 g_{\mu\nu}R$ can be moved to the right-hand side giving the so-called *trace-reversed Einstein equations*

$$R_{\mu\nu} = -8\pi G\left[T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right]\,. \tag{8.7}$$

[72] The sign in the Einstein Equation depends on convention. We follow Weinberg's notation in this section (different from Dodelson's).

The advantage of this form of the equations is that the perturbations to $T_{\mu\nu}$ are somewhat easier to compute than those to $R_{\mu\nu}$. The disadvantage is that it is not at all obvious that taking the covariant derivative gives the conservation of $T_{\mu\nu}$ via the contracted Bianchi identities.
Linearising these equations is lengthy but straightforward. We leave it as an exercise. Nowadays the calculation can be done in less than a second using publicly available codes such as the mathematica package xPand [50] (which uses xAct). We discuss this in Prob. P.8.1. Computing $\delta R_{\mu\nu}$ and $\delta T_{\mu\nu}$ and substituting it into

$$\delta R_{\mu\nu} = -8\pi G\left[\delta T_{\mu\nu} - \frac{1}{2}\bar{g}_{\mu\nu}\delta T^{\lambda}_{\lambda} - \frac{1}{2}h_{\mu\nu}\bar{T}^{\lambda}_{\lambda}\right] , \tag{8.8}$$

gives the scary looking expressions

$$\begin{aligned}
-\frac{1}{M_{\rm Pl}^2}\left(\delta T_{ij} - \frac{a^2}{2}\delta_{ij}\delta T^{\lambda}_{\lambda}\right) &= -\frac{1}{2}\partial_{ij}h_{00} - \delta_{ij}\left[\left(2\dot{a}^2 + a\ddot{a}\right)h_{00} + \frac{1}{2}a\dot{a}\dot{h}_{00}\right] + \left(H^2 + 3\frac{\ddot{a}}{a}\right)h_{ij} \\
&\quad + \frac{1}{2a^2}\left(\partial_l^2 h_{ij} - \partial_{l(i}h_{j)l} + \partial_{ij}h_{ll}\right) + \frac{H}{2}\left(\dot{h}_{ij} - \delta_{ij}\dot{h}_{ll}\right) \\
&\quad - \frac{1}{2}\ddot{h}_{ij} + H\delta_{ij}\left(Hh_{ii} + \partial_l h_{0l}\right) + \frac{1}{2}\left(\partial_{(i}\dot{h}_{j)0} + H\partial_{(i}h_{j)0}\right) ,
\end{aligned} \tag{8.9}$$

$$\begin{aligned}
-\frac{1}{M_{\rm Pl}^2}\delta T_{j0} &= H\partial_j h_{00} + \frac{1}{2a^2}\left(\partial_l^2 h_{j0} - \partial_{jl}h_{l0}\right) + \left(H^2 + 2\frac{\ddot{a}}{a}\right)h_{j0} \\
&\quad + \frac{1}{2}\partial_t\left[\frac{1}{a^2}\left(\partial_j h_{kk} - \partial_k h_{jk}\right)\right] ,
\end{aligned} \tag{8.10}$$

$$\begin{aligned}
-\frac{1}{M_{\rm Pl}^2}\left(\delta T_{00} + \frac{1}{2}\delta T^{\lambda}_{\lambda}\right) &= \frac{1}{2a^2}\partial_l^2 h_{00} + \frac{3}{2}H\dot{h}_{00} - \frac{1}{a^2}\partial_i\dot{h}_{i0} + 3\left(H^2 + \frac{\ddot{a}}{a}\right)h_{00} \\
&\quad + \frac{1}{2a^2}\left[\ddot{h}_{ii} - 2H\dot{h}_{ii} + 2\left(H^2 - \frac{\ddot{a}}{a}\right)h_{ii}\right] ,
\end{aligned} \tag{8.11}$$

where $\partial_l^2 \equiv \partial_l\delta^{lk}\partial_k$ and (see Footnote **??**)

$$\delta T^{\mu}{}_{\nu} = \bar{g}^{\mu\lambda}\left[\delta T_{\lambda\nu} - h_{\lambda\rho}\bar{T}^{\rho}{}_{\nu}\right] . \tag{8.12}$$

Notice that, as implied by the Bianchi identities [73] , the four metric perturbations $h_{\mu0}$ appear with at most one time derivative in these equations and are therefore non-dynamical. It is useful to discuss this quantitatively in terms how many initial conditions we need to specify to find a solution. Heuristically we have ten components for the symmetric tensor $h_{\mu\nu}$ minus four gauge transformations and four constraints leading to two propagating degrees of freedom[74], namely the two helicities of the graviton.

[73]To see this recall that the Bianchi identities (A.33), which are off-shell identities valid irrespectively of the equations of motion, say $\nabla^{\mu}G_{\mu\nu}$ with $G_{\mu\nu}$ the Einstein tensor on the left-hand side of EE's. Expanding the covariant derivative we find

$$\partial_t G^{t\beta} = -\partial_k G^{k\beta} - \Gamma^{\alpha}_{\alpha\gamma}G^{\beta\gamma} - \Gamma^{\beta}_{\alpha\gamma}G^{\alpha\gamma} . \tag{8.13}$$

Since the right-hand side has at most second derivatives of the metric (in $G_{\mu\nu}$), we conclude that $G^{t\beta}$ has at most first derivatives. It takes a bit more work and the ADM formalism to specify which components of the metric appear with at most one derivative. At linear order, by inspection we see that it is $h_{\mu0}$.

[74]More precisely we can use the Hamiltonian formulation of general relativity. For the ten fields $h_{\mu\nu}$, $N_f = 10$, one finds eight first class constraints, $N_1 = 8$, and zero second class constraints, $N_2 = 0$, giving $N_f - N_1 - 2N_2 = 2$ degrees of freedom.

The linearised conservation of the energy momentum tensor takes the following form

$$\delta\left(\nabla_\mu T^\mu{}_\nu\right) = \partial_\mu \delta T^\mu{}_\nu + \bar\Gamma^\mu_{\mu\lambda}\delta T^\lambda{}_\nu - \bar\Gamma^\lambda_{\mu\nu}\delta T^\mu{}_\lambda + \delta\Gamma^\mu_{\mu\lambda}\bar T^\lambda{}_\nu - \delta\Gamma^\lambda_{\mu\nu}\bar T^\mu{}_\lambda = 0\,. \tag{8.14}$$

Using

$$\delta\Gamma^0_{0i} = H h_{0i} - \frac{1}{2}\partial_i h_{00}\,, \tag{8.15}$$

this can be written as

$$\partial_0 \delta T^0{}_j + \partial_i \delta T^i{}_j + 2H\delta T^0{}_j - a^2 H \delta T^j{}_0 - (\bar\rho + \bar p)\left(\frac{1}{2}\partial_j h_{00} - H h_{j0}\right) = 0 \tag{8.16}$$

$$\partial_0 \delta T^0{}_0 + \partial_i \delta T^i{}_0 + 3H\delta T^0{}_0 - H\delta T^i{}_i + \frac{\bar\rho + \bar p}{a^2}\left(H h_{ii} - \frac{1}{2}\dot h_{ii}\right) = 0\,. \tag{8.17}$$

We would like to parameterise the perturbations $h_{\mu\nu}$ and $\delta T_{\mu\nu}$ in such a way as to simplify the calculation as much as possible. Experience teaches us that it is wise to choose variables that transform nicely under the symmetry of the system. First, because the theories we are working with are fully covariant, we can choose coordinates that simplify the calculation. Second, the background around which we are expanding is invariant under rotations and translations. By rotating and translating a perturbation that solves the equations of motion, we obtain another, in general different perturbation that also solves them. This is a linear operation and so, in more mathematical terms, the space of solutions provides a linear *representation* of the isometry group $SO(3) \times \mathbb{R}^3 = ISO(3)$, called the Euclidean group. The building blocks of these representations are irreducible representations (irreps). In this context, an irrep is a family of solutions that can all be transformed into each other by some element of ISO(3). Intuitively[75] these can be thought of as cosmological "particles", the building blocks of more general cosmological perturbations. The construction of irreps of the non-compact group ISO(3) is easily performed using the method of "induced representations". The idea is to find a representation for a subgroup, in this case the *little group*, and extend that representation to the whole group. A summary of the derivation in Sec. 8.7 is that perturbations are classified by the norm of their Fourier three momentum $\mathbf{k}^2$ and by their helicity, namely the angular momentum in their direction of motion $p^i J_i$, with eigenvalues $0, \pm1, \pm2, \ldots$. A convenient basis for each irrep is characterised by the Fourier three momentum $\mathbf{k}$ and the helicity. The main advantage of the above choice of variables is that, to linear order, perturbations with different momentum and helicity do not interact with each other.

In summary, three observations make the task of solving the above equations more manageable:

- *Fourier decomposition*: because we expand around a *homogeneous* background, different Fourier modes decouple from each other at linear order.
- *Scalar-Vector-Tensor (SVT) decomposition*: because we work with general covariant theories and we expand around an *isotropic* background, objects that transform differently under spatial rotations do not mix with each other at linear order.

[75] This discussion follows closely the analogous introduction of particles in relativistic QFT. In relativistic theories, particles are the irreducible representation of the Poincaré group. These are first classified by their mass $m^2 = -p^\mu p_\mu$. Then, for massive particles $m > 0$, they are further classified by their spin, i.e. the eigenvalues of the total spin operator J^2 and the spin in one of the three spatial directions J_z. Massless particles are instead classified by their helicity, i.e. the eigenvalue of their angular momentum in their direction of motion $p^i J_i$.

- *Gauge transformations*: since we are dealing with GR, a covariant theory of gravity, there is some redundancy in our description due to the arbitrary choice of coordinates. One can always perform a coordinate transformation, which we will soon interpret as a gauge transformation on the fields, to conveniently simplify the equations.

Notice that the first two simplifications crucially rely on working at *linear order*, while the last survives at all orders in perturbation theory. Let us discuss these three points in detail.

8.2 Fourier and scalar-vector-tensor decomposition

Let us now see how the isometries restrict the possible interaction among these perturbations.

Fourier decomposition We claim that, because of the homogeneity of the background, different Fourier modes decouple from each other at linear order. To see why this is the case, consider the general form of the linearized equations of motion

$$\sum_A \mathcal{O}_A \, \mathrm{Pert}_A(\mathbf{x}, t) = 0 \,, \tag{8.18}$$

where A enumerates all perturbations $\mathrm{Pert}_A = \{h_{\mu\nu}, \delta T_{\mu\nu}\}$ and $\mathcal{O}_A$ are linear differential operators acting on the perturbations (e.g. $H(t)\partial_t$ or $a^{-2}\partial_i\partial_i$). Because of general covariance, these operators must be constructed out of covariant spacetime derivates ∇_μ and other tensorial objects evaluated on the background

$$\mathcal{O}_A = \mathcal{O}_A(\nabla_\mu, \bar{g}_{\mu\nu}, \bar{T}_{\mu\nu}) = \mathcal{O}_A(\partial_\mu, \partial_t^{\#}\bar{g}_{\mu\nu}(t), \partial_t^{\#}\bar{T}_{\mu\nu}(t)) \,. \tag{8.19}$$

Since the background is homogeneous, $\mathcal{O}_A$ cannot depend on space $\mathbf{x}$, but it does in general depend on time through $\bar{g}_{\mu\nu}(t)$ and $\bar{\delta}T_{\mu\nu}(t)$. As we take the Fourier transform of (8.18), we find

$$\int d^3x e^{-i\mathbf{x}\mathbf{k}} \sum_A \mathcal{O}_A \, \mathrm{Pert}_A(\mathbf{x}, t) = \sum_A \tilde{\mathcal{O}}_A \, \mathrm{Pert}_A(\mathbf{k}, t) = 0 \,, \tag{8.20}$$

with (see Eq. (12.9) for my Fourier conventions)

$$\mathrm{Pert}_A(\vec{k}, t) = \int d^3x \, e^{-i\mathbf{x}\mathbf{k}} \, \mathrm{Pert}_A(\mathbf{x}, t) \tag{8.21}$$

$$\text{and} \quad \tilde{\mathcal{O}}_A = \mathcal{O}_A(\partial_t, \partial_i \to ik_i, \partial_t^{\#}\bar{g}_{\mu\nu}(t), \partial_t^{\#}\bar{T}_{\mu\nu}(t)) \tag{8.22}$$

where all spatial derivatives have been integrated by part to act on $e^{-i\mathbf{x}\mathbf{k}}$ hence giving $i\mathbf{k}$. While Eq. (8.18) was a *partial* differential equation, Eq. (8.20) is now a infinite set of *ordinary* differential equations, one for each $\mathbf{k}$. Since in each equation only one $\mathbf{k}$ appears in the arguments of Pert_A, different Fourier modes with $\vec{k} \neq \vec{k}'$, decouple from each other. In other words, at linear order one can always look for solutions with a single, monochromatic perturbation with wavevector $\mathbf{k}$ in an otherwise unperturbed background universe. Any linear combination of these solutions is also a solution (linear superposition). Finally, notice that k is the Fourier conjugate of x, and so it is a *comoving* momentum. Physical momentum is instead $k_{\mathrm{phy}} = k/a$, in the same way that $x_{\mathrm{phy}} = xa$.

Scalar-vector-vensor decomposition Let us know take advantage of the isotropy of the background. Because we work with general covariant theories and we expand around an isotropic background, perturbations that transform differently under *spatial rotations* do not mix with each other at linear order. Let us see why.

Rotations are changes of coordinates of the form

$$\{x^0, x^i\} \to \{x'^0, x'^{i'}\} = \{x^0, R^{i'}{}_i x^i\} \quad \Rightarrow \quad J^{\mu'}{}_\mu \equiv \frac{\partial x'^{\mu'}}{\partial x^\mu} = \begin{pmatrix} 1 & \\ & R^{i'}{}_i \end{pmatrix} \tag{8.23}$$

and the Jacobian $J^{\mu'}{}_\mu$ has only non-trivial spacial components. Let us use this to compute how different objects transform. Consider the simplest objects, namely diff-scalars and their perturbations. [76] Some examples are the density ρ, the pressure p, the number density n or the temperature T. For general changes of coordinates $x^\mu \to x'^\mu(x)$, any diff-scalar s transforms as

$$s'(x') = s(x(x'))\,,$$

where on the right-hand side $x(x')$ is the solution of $x' = x'(x)$. Diff-vectors, such as u^μ, and symmetric tensors, such as $g_{\mu\nu}$, have a richer structure. From their transformation properties under general diffs, (A.4), it is immediate to see that when all indices are in the time direction, these objects transform as rotation-scalars, e.g.

$$u^0(x) \to u'^0(x') = J^0{}_\mu u^\mu(x) = u^0(x)\,, \quad h_{00}(x) \to h'_{00}(x') = J_0{}^\mu J_0{}^\nu h_{\mu\nu}(x) = h_{00}(x)\,. \tag{8.24}$$

When the indices are in the spatial directions we can apply the Hodge decomposition, which is a generalization of Helmholtz decomposition familiar from electromagnetism. For example, any spatial vector v^i can be decomposed as[77]

$$v_i = \omega_i + \partial_i \theta\,, \tag{8.25}$$

where ω_i is divergence-free or *transverse*, namely $\partial_i \omega_i = 0$. To find θ, we have to solve the differential equation $\nabla^2\theta = \partial^i v_i$. On a topologically trivial space such as $\mathbb{R}^3$ and assuming that v_i vanishes at spatial infinity, this Poisson equation can be uniquely solved for θ. Then in turn we have $\omega_i = v_i - \partial_i\theta$.
The Helmholtz decomposition is covariant under rotations if we assume that θ transform as a rotation-scalar (see P.8.2) and ω_i as a rotation-vector, i.e.

$$\omega'_{i'}(x', t) = R_{i'}{}^i \omega_i(x, t)\,. \tag{8.26}$$

[76] Some nomenclature. The terms scalar, vectors and tensor may refer to the transformation of an object either under general change of coordinates, a.k.a. *diffeomorphisms* (diffs), or only under spatial rotations. To be crystal clear, in this section we'll denote these two concepts differently. We define diff-scalars, diff-vectors and diff-tensors objects that transform covariantly under general changes of coordinates, as in (A.4). Analogously, rotations-scalars, rotation-vectors and rotation-tensors will be objects that transform appropriately under rotation, as we will see shortly. In the rest of the lectures instead the difference will hopefully be clear from the context.

[77] Helmholtz theorem states that any smooth vector field that is rapidly decreasing at infinity can be uniquely decomposed into a curl-free vector and a divergence-free vector. In $\mathbb{R}^3$, these vectors can be written as the gradient of a scalar potential, e.g. the electro-static potential, and the curl of a vector potential, e.g. the vector potential generating a magnetic field. In cosmology it is customary to work with the scalar potential, e.g. θ in Eq. (8.25), and the divergence-free vector, e.g. ω_i in Eq. (8.25).

Exactly the same Helmholz decomposition can be used for any tensor with one spatial and one time index such as h_{0i}. The last object we will need to decompose is the spatial-spatial part h_{ij}. It is straightforward to see that the trace $h^i{}_i$ taken with the background metric is a rotation-scalar. For the remaining 5 component we can use a generalisation of Helmholz decomposition to any tensor, which breaks up h_{ij} into two rotation-scalars, one transverse vector and a *transverse-traceless tensor*. Let us introduce some notation to conveniently deal with the SVT decomposition. The metric perturbation $h_{\mu\nu}$ is a symmetric 4×4 matrix with 10 independent entries. They can be SVT-decomposed as follows[78]

$$\begin{aligned} h_{00} &= -E\,, \\ h_{i0} &= a\left[\partial_i F + G_i\right]\,, \\ h_{ij} &= a^2\left[\delta_{ij}A + \partial_{ij}B + \partial_{(i}C_{j)} + D_{ij}\right]\,, \end{aligned} \tag{8.27}$$

with the transverse vectors C_i and G_i and the transverse-traceless tensor D_{ij} satisfying

$$\partial_i G_i = \partial_i C_i = D_{ii} = \partial_i D_{ij} = 0\,.$$

In P.8.3 you will explicitly perform this decomposition. We have four scalars E, A, B and F, two transverse vectors C_i and G_i, with two "polarizations" each, and one transverse traceless tensor D_{ij}, also with two polarizations. These add up to 10, as expected.

Fluids For $T_{\mu\nu}$ we start with a perfect fluid with normalized velocity $u_\mu u^\mu = -1$,

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu} = \rho u_\mu u_\nu + p(g_{\mu\nu} + u_\mu u_\nu) \tag{8.28}$$

Here $(g_{\mu\nu} + u_\mu u_\nu)$ projects onto the spatial hypersurface perpendicular to u^μ since

$$u^\mu(g_{\mu\nu} + u_\mu u_\nu) = u_\nu(1 + u_\mu u^\mu) = 0\,. \tag{8.29}$$

Since $T_{\mu\nu}$ is written in terms of two scalars p and ρ and a four-vector u_μ, its SVT decomposition is

$$\rho = \bar\rho + \delta\rho\,, \qquad p = \bar p + \delta p\,, \qquad u_\mu = \bar u_\mu + \delta u_\mu\,, \tag{8.30}$$

$$\bar u_\mu = (-1, 0, 0, 0)\,, \qquad \delta u_\mu = (\delta u_0, \partial_i \delta u + \delta u_i^V)\,. \tag{8.31}$$

To maintain the normalization of u^μ, one needs at linear order

$$-1 = u_\mu g^{\mu\nu} u_\nu = (-1 + \delta u_0, \delta u_i)_\mu (\bar g^{\mu\nu} + h^{\mu\nu})(-1 + \delta u_0, \delta u_i)_\nu = -1 + 2\delta u_0 + h^{00}\,, \tag{8.32}$$

and so $\delta u_0 = -h^{00}/2 = h_{00}/2$. Because we restricted ourselves to perfect fluids, our parameterization so far contains only 5 components, three scalars and one transverse vector. To account for all 10 independent components of a generic $T_{\mu\nu}$ we generalized this SVT-decomposition as follows:

$$\begin{aligned} \delta T_{00} &= -\bar\rho h_{00} + \delta\rho\,, \\ \delta T_{i0} &= \bar p h_{0i} - (\bar\rho + \bar p)\left[\partial_i \delta u + \delta u_i^V\right]\,, \\ \delta T_{ij} &= \bar p h_{ij} + a^2\left[\delta_{ij}\delta p + \partial_{ij}\pi^S_{ij} + \partial_{(i}\pi^V_{j)} + \pi^T_{ij}\right]\,, \end{aligned} \tag{8.33}$$

[78]The factors of a in these definitions are of course arbitrary and chosen for future convenience.

Here π^S, π^V and π^T are called *anisotropic inertia* and are a property of a given fluid that needs to be specified to close the system of equations. For example, all anisotropic inertia vanishes for a perfect fluid, as can be seen by comparing to Eq. (1.34). Anisotropic inertia generate dissipation and capture effects such as shear and bulk viscosity. In summary, a generic dissipative fluid has four scalars ($\delta\rho$, δp, δu and π^S), two transverse vectors (π^V and δu^V) and one transverse traceless tensor (π^T), adding up again to 10.

Now the essential point. Since the theory and the FLRW background are invariant under rotations, the equations of motion must be covariant under rotations, i.e. invariant in form. This means that to linear order two fields can appear in the same equation only if they have the same transformations under rotations. We hence see that rotation-scalars, transverse vectors and transverse-traceless tensors decouple from each other at linear order. Decoupling means that in solving the equations of motion for one of the three types of perturbations, we can set the others to zero. Any combination of the three sets of scalar, vector and tensor solutions is also a solution. To see decoupling in an intuitive way, note that it is impossible to construct a non-vanishing scalar from a transverse vector ω_i or a transverse-traceless tensor v_{ij} using only derivatives and background quantities at linear order. In fact, the only object that one can use to contract the spatial indices are the background spatial metric, proportional to δ_{ij}, and spatial derivatives ∂_i. Any contraction of all indices is identically zero because of the transverse and traceless conditions. A similar argument shows that all other mixing terms must vanish.

8.3 Gauge transformations

Since we are dealing with GR, one can always perform a coordinate transformation to simplify the equations. Consider the coordinate transformation

$$x^\mu \to x'^\mu = x^\mu + \epsilon^\mu(x)\,, \tag{8.34}$$

for arbitrary $\epsilon^\mu(x)$. We will be interested in transformations that make some perturbations vanish identically, so we will restrict ourselves to cases in which ϵ^μ is a regular and decreasing function at spatial infinity and it is of first order in perturbations $\epsilon \sim \mathcal{O}(h_{\mu\nu}, \delta T_{\mu\nu})$. While we know that tensors such as $g_{\mu\nu}$ and $T_{\mu\nu}$ transform as in Eq. (A.4), we have now the additional complication that every tensor is split between a background and a perturbation, as e.g. in Eq. (8.1). We have therefore an ambiguity on how the background and the perturbation transform separately, while keeping the covariance of the full tensor. A convenient an very common way to solve this ambiguity is to work with so called *gauge transformations*, in which case the background is kept fixed and all the transformations of the full tensor are attributed to the perturbations. More in detail, the rules are the following

1. Transform the full tensor covariantly, as in Eq. (A.4), but keep the background unchanged
2. Attribute all the transformation to the perturbations

For example, the perturbation δs to a scalar field transforms as

$$\delta s(x) \to \delta s'(x) = s'(x) - \bar{s}'(x) = s(x-\epsilon) - \bar{s}(s) = \delta s(x) - \epsilon^\mu \partial_\mu(\bar{s}(x) + \delta s(s))\,, \tag{8.35}$$

where we used that background does not change, $\bar{s}'(x) = \bar{s}(s)$ and that s changes as a scalar $s'(x) = s(x - \epsilon)$ for $x \to x' = x + \epsilon$. This tells us that the transformation $\Delta\delta s$ of a scalar field perturbation $\delta s(x) = s(x) - \bar{s}(x)$ under the change of coordinates $x \to x' + \epsilon$ is

$$\Delta\delta s(x) \equiv \delta s'(x) - \delta s(x) = -\epsilon^\mu \partial_\mu s(x) + \mathcal{O}(\epsilon^2)\,. \tag{8.36}$$

Since we will always work with a homogeneous background, $\bar{s}(x) = \bar{s}(t)$, this simplifies to

$$\Delta\delta s = -\epsilon^0 \dot{\bar{s}} + \mathcal{O}(\epsilon^2, \epsilon h_{\mu\nu}, \epsilon\delta T_{\mu\nu})\,. \tag{8.37}$$

The same rules apply to vectors and symmetric two-tensors, for which one finds (see P.8.4)

$$\Delta\delta V^\mu \equiv V'^\mu(x) - V^\mu(x) = -\epsilon^\nu\partial_\nu V^\mu + V^\nu\partial_\nu\epsilon^\mu = -\epsilon^\nu\nabla_\nu V^\mu + V^\nu\nabla_\nu\epsilon^\mu\,, \tag{8.38}$$

$$\Delta h_{\mu\nu}(x) \equiv g'_{\mu\nu}(x) - g_{\mu\nu}(x) = -\nabla_\mu\epsilon_\nu - \nabla_\nu\epsilon_\mu\,. \tag{8.39}$$

The attentive reader will have noticed that these gauge transformations are nothing but Lie derivatives. In general, the transformation of the perturbations to any tensor are given by minus its Lie derivative in the direction ϵ. At linear order this simplifies to

$$\Delta\delta\text{Tensor} = -\mathcal{L}_\epsilon\text{Tensor} = -\mathcal{L}_\epsilon\overline{\text{Tensor}} + \mathcal{O}(\epsilon^2)\,, \tag{8.40}$$

where in the last equality we used the fact that ϵ is already first order in perturbation. In particular, notice that all covariant derivatives in (8.38) should be computed from the background metric. So far we have discussed gauge transformation for the full diffeomorphism tensors, but we would like to know how each SVT component transforms. To this end, we SVT-decompose the gauge parameter ϵ^μ by defining

$$\epsilon_\mu = \{\epsilon_0, \partial_i\epsilon^S + \epsilon_i^V\}\,, \tag{8.41}$$

with ϵ_0 and ϵ^S two rotation-scalars and ϵ_i^V a transverse vector, $\partial_i\epsilon_i^V = 0$. Using Eq. (8.40) and the SVT decomposition Eq. (8.27), we find the following linear gauge transformations of the SVT components for the metric

$$\begin{aligned}
&\Delta A = 2H\epsilon_0\,, && \Delta B = -\frac{2}{a^2}\epsilon^S\,, && \\
&\Delta F = \frac{1}{a}\left(-\epsilon_0 - \dot{\epsilon}^S + 2H\epsilon^S\right)\,, && \Delta E = 2\dot{\epsilon}_0\,, && \\
&\Delta C_i = -\frac{1}{a^2}\epsilon_i^V\,, && \Delta G_i = \frac{1}{a}\left(-\dot{\epsilon}_i^V + 2H\epsilon_i^V\right) && \Delta D_{ij} = 0\,,\,,
\end{aligned} \tag{8.42}$$

Note that the transformations of the scalar perturbations only involve the scalar parts of the gauge parameter ϵ_0 and ϵ^S, while those of the vector perturbations the vector part ϵ_i^V, as expected. Since there is no traceless-transverse part of ϵ_μ we find that the traceless-transverse part of the metric D_{ij} is gauge invariant to linear order. To compute the gauge transformations of scalar quantities such as ρ or p we need the gauge parameter with upper indices $\epsilon^\mu = g^{\mu\nu}\epsilon_\nu \sim \bar{g}^{\mu\nu}\epsilon_\nu$. Using the SVT decomposition Eq. (8.33) for the energy-momentum tensor we have

$$\begin{gathered}
\Delta\delta\rho = \dot{\bar{\rho}}\epsilon_0\,, \quad \Delta\delta p = \dot{\bar{p}}\epsilon_0 \quad \Delta\delta u = -\epsilon_0\,, \\
\Delta\pi^S = \Delta\pi_i^V = \Delta\pi_{ij}^T = \Delta\delta u_i^V = 0\,,
\end{gathered} \tag{8.43}$$

where we used $\epsilon_0 = -\epsilon^0$ and $\epsilon_i = a^2\epsilon^i$. Notice that the transformations of the perturbations $\delta\rho$ and δp to diff-scalars are obtained from Eq. (8.37), and those of δu^μ from Eq. (8.38). It comforting to note that the condition for a perfect fluid $\pi^{S,V,T} = 0$ and for an irrotational fluid $u_i^V = 0$ are gauge invariant as expected.

8.4 Vector and tensor perturbations

Because we work only with diff-invariant theories, all equations of motions can be written as the vanishing of some covariant tensor. For example, let's focus on the Einstein equations, i.e. $M_{\rm Pl}^2 G_{\mu\nu} + T_{\mu\nu} = 0$. We can then apply the same SVT decomposition to this 2-index tensor and extract four scalar equations, two transverse-vector equations and one transverse-traceless-tensor equation. We will start with the vector and tensor equations since they are the simplest to study.

Vector perturbations

Vectors[79] decay with time and so do not play much of a role in cosmology[80]. To see this, let us take advantage of the SVT decomposition and set all scalar and tensor perturbations to zero. We are left with

$$\begin{aligned} h_{00} &= 0\,, & \delta T_{00} &= 0\,, \\ h_{0i} &= aG_i\,, & \delta T_{0i} &= \bar p a G_i - (\bar\rho + \bar p)\,\delta u_i^V \\ h_{ij} &= a^2 \partial_{(i} C_{j)}\,, & \delta T_{ij} &= a^2\left(\bar p \partial_{(i} C_{j)} + \partial_{(i}\pi^V_{j)}\right)\,, \end{aligned} \tag{8.44}$$

Plugging this into the linearized momentum conservation equation $T^{i\mu}{}_{;\mu} = 0$, (8.16), one finds

$$\partial^2 \pi_j^V + \partial_0\left[(\bar\rho + \bar p)\,\delta u_j^V\right] + 3H\,(\bar\rho + \bar p)\,\delta u_j^V = 0\,. \tag{8.45}$$

All ingredients of the standard cosmological model, namely baryons, dark matter, dark energy, photons and neutrinos, behave as a perfect fluids to good approximation and so we neglect the anisotropic inertia[81]. We then find

$$(\bar\rho + \bar p)\,\delta u_j^V \simeq a^{-3}\,. \tag{8.46}$$

Using (8.44) into the linearized $0i$ part of the trace-reversed Einstein equations, (8.10), one finds

$$8\pi G\,(\bar\rho + \bar p)\,\delta u_j^V a = \frac{1}{2}\partial^2\left(G_j - a\dot C_j\right)\,, \tag{8.47}$$

and so $G_i - \dot C$ decays as a^{-2} by virtue of (8.46). Using Eq. (8.42), one can prove that this combination is indeed the only gauge-invariant vector mode (see Prob. P.8.5).

Tensor perturbations

From the space-space (ij) components of the EE's one can extract the transverse traceless part following P.8.3. But given that we proved that SVT components decouple, it is much easier to set all scalars and vectors to zero and keep only D_{ij} in the linearised EE's. Substituting

$$h_{0\mu} = 0 = \delta T_{0\mu}\,, \qquad h_{ij} = a^2 D_{ij} \quad \text{and} \quad \delta T_{ij} = a^2\left(\bar p D_{ij} + \pi^T_{ij}\right) \tag{8.48}$$

[79] Henceforth we simply write "vectors" and "tensors" omitting "transverse" and "transverse traceless".
[80] An exception are the speculated primordial magnetic fields (see [62] for a review).
[81] Neutrinos have some anisotropic inertia as they become non-relativistic, and this results in a 10% correction to the spectrum of tensor modes [66]

into (8.9) and recalling that $D_{ii} = \partial_i D_{ij} = 0$, one finds (see **??**)

$$\ddot{D}_{ij} + 3H\dot{D}_{ij} - \frac{\partial^2}{a^2} D_{ij} = 8\pi G \pi^T_{ij} \,. \tag{8.49}$$

This equation contains almost[82] all linear terms in D_{ij} and its derivatives that are allowed by symmetries of the problem up to two derivatives. The tensor anisotropic inertia π^T is small for all components of the universe. The largest contributors are neutrinos and their π^T eventually leads to a 20% reduction in the amplitude of D_{ij} (see [66] or Sec 6.6 of [68] for a detailed calculation). After neglecting π^T, (8.49) takes the same form as the equation of motion for a massless scalar field in FLRW (see P.8.6). The solution is best understood in Fourier space

$$\ddot{D}_{ij} + 3H\dot{D}_{ij} + \frac{k^2}{a^2} D_{ij} = 0 \,. \tag{8.50}$$

Because of parity invariance of the background and of GR, each of the two independent components of D_{ij} have the same time dependence. To make this more explicit, let us separate the index structure from the time dependence:

$$D_{ij}(t, \mathbf{k}) = \sum_s \epsilon^s_{ij}(\mathbf{k}) \mathcal{D}_s(t, k) \,. \tag{8.51}$$

where $\epsilon^s_{ij}(\mathbf{k})$ are *polarization tensors* and s takes two values corresponding to the two polarizations. There are many possible choices of basis for the two polarization tensors. A common choice in the study of gravitational waves at interferometers are the “plus” and “cross” polarizations, $s = +, \times$. These polarization are somewhat equivalent to linear polarization in electromagnetism. Another common choice inherited from particle physics is circular polarization, $s = \pm 2$, where each polarization is an eigenvector of rotations around the momentum direction. We will not need to make any specific choice here, but explicit expressions are derived in Appendix **??**. In general, polarization tensors are complex and satisfy

$$\epsilon^s_{ii}(\mathbf{k}) = k^i \epsilon^s_{ij}(\mathbf{k}) = 0 \qquad \text{(transverse and traceless)}\,, \tag{8.52}$$

$$\epsilon^s_{ij}(\mathbf{k}) = \epsilon^s_{ji}(\mathbf{k}) \qquad \text{(symmetric)}\,, \tag{8.53}$$

$$\epsilon^s_{ij}(\mathbf{k}) \epsilon^s_{jk}(\mathbf{k}) = 0 \qquad \text{(lightlike)}\,, \tag{8.54}$$

$$\epsilon^s_{ij}(\mathbf{k}) \epsilon^{s'}_{ij}(\mathbf{k})^* = 2\delta_{ss'} \qquad \text{(normalization)}\,, \tag{8.55}$$

$$\epsilon_{ij}(\mathbf{k})^* = \epsilon_{ij}(-\mathbf{k}) \qquad (\gamma_{ij}(x) \text{ is real})\,. \tag{8.56}$$

Now we are interested in solving

$$\ddot{\mathcal{D}}_s(t, k) + 3H\dot{\mathcal{D}}_s(t, k) + \frac{k^2}{a^2} \mathcal{D}_s(t, k) = 0 \,. \tag{8.57}$$

In practical applications, one usually specifies the expansion history $a(t)$, and then solves this equation numerically to the required precision. Here we will instead look at some approximate solutions valid for any cosmology, i.e. for arbitrary $a(t)$. First, consider the superHubble regime

[82] The additional parity-odd term $\epsilon_{ik(l} \partial_k D_{lj)}$ would violate the equivalence principle, but $\epsilon_{ik(l} \dot{\partial}_k D_{lj)}$ could appear in dissipative situations [56].

in which the wavelength of the perturbation is much larger than the Hubble radius $k/a \ll H$. Then we can drop the spatial derivatives[83]

$$k \ll aH: \qquad \ddot{\mathcal{D}}_s + 3H\dot{\mathcal{D}}_s \simeq 0 \quad \Rightarrow \quad \dot{\mathcal{D}}_s \propto a^{-3}, \tag{8.58}$$

and so the two independent superHubble solutions are

$$\mathcal{D}_s(t, k \ll aH) = \left[A_s(k) + B_s(k)a^{3(w-1)/2}\right], \tag{8.59}$$

The time dependent solution is decaying for $w < 1$, which is always satisfied in standard cosmologies, and so it can usually be neglected after some efoldings of superHubble evolution. In the opposite regime of subHubble perturbations $k \gg aH$, we can solve (8.49) in the WKB approximation. By making an Ansatz $\mathcal{D}_s = X(t)\exp\left[\pm ik\int^t dt'/a(t')\right]$ and solving the resulting differential equation for $X(t)$ to leading order in $k \gg aH$, one finds $X \propto a^{-1}$. So, the two independent subHubble solutions are

$$\mathcal{D}_s(t,k) = \frac{\tilde{A}_s\cos(k\tau(t)) + \tilde{B}_s\sin(k\tau(t))}{a} \qquad (k \gg aH), \tag{8.60}$$

with $\tau \equiv \int^t dt'/a(t')$. These solutions describe the oscillations of gravitational waves as they propagate, but we also notice that the amplitudes decay as a^{-1} due to the expansion of the universe. Notice that, if in a non-standard cosmology parity is broken, the two polarizations could have different initial conditions and different dynamics.

8.5 Scalar perturbations

It is time to tackle the most complicated and most relevant modes for cosmology: scalar perturbations. Let us start with a simple counting, assuming for simplicity that we have only one fluid[84]. We have four independent scalar equations (00, ii, longitudinal $0i$, and longitudinal ij parts of the Einstein Equations[85]) for 8 variables (four in the metric, A, B, E and F, and four in δT, namely $\delta\rho, \delta p, \delta u$ and π^S). The pressure p and anisotropic stresses $\pi^{S,V,T}$ depend on the property of the fluid under consideration and need to be specified by some constitutive equations, such as the equation of state $p = p(\rho, \dots)$. For example, for a relativistic perfect fluid $p = \rho/3$, while for a non-relativistic one $0 =\simeq p \ll \rho$. Also, for a perfect fluid, which is a good approximation in most cosmological applications, the anisotropic inertia vanish $\pi^{S,V,T} = 0$. This determines[86] two scalars, namely π^S and δp. We are still left with 6 variables for 4 equation, but we have not used the two scalar gauge transformations ϵ^0 and ϵ^S. One can now proceed in two ways. One can work only with gauge-invariant combinations, namely 4 independent linear combinations of the 6 scalars that are invariant

[83]This is sometimes called the "separate universe" approximation because after dropping the spatial derivatives every superHubble patch of the universe evolves completely independently from the others. One can also keep subleading order in spatial derivatives.

[84]In practice there will be several different components of the universe. Some components might be interacting with each other such as electron, baryons and photons before recombination, while some components might be decoupled, such as neutrinos at $T \ll$MeV. The energy momentum tensor is separately conserved for each set of mutually interacting components.

[85]Notice that the equations for the conservation of the energy-momentum tensor are *not* independent

[86]Given a simple equation of state $p = p(\rho)$, one finds $\delta p = (\partial p/\partial\rho)\,\delta\rho$.

under the gauge transformations (8.42) and (8.43). We will encounter two such variable later, (12.61). Alternatively, one can *fix the gauge* and work with a particular set of coordinates. This second approach is somewhat more convenient and will be followed in this course.
The idea of fixing the gauge is to choose coordinates that correspond to the constant hypersurfaces of some of the perturbations, so that those perturbations appear constant. In other words, we can choose ϵ^0 and ϵ^S in Eq. (8.41) in such a way to cancel whatever profile of some of the scalar perturbations, using the transformation properties in (8.42). Since there are 6 scalar perturbations but only two scalar gauge parameters, there are clearly many different possible choices (in fact infinitely many). Notice that the gauge parameters ϵ^μ need to vanish at spatial infinity in the same way as the physical perturbations they need to cancel. In this sense these are *small gauge transformations*. See below Eq. (8.95) for a discussion of large gauge transformations. Let us see the most commonly used gauge choices

Newtonian gauge Using (8.42), we see that

$$\begin{cases} \epsilon^S = a^2 B/2 \\ \epsilon_0 = aF - \frac{a^2}{2}\dot{B} \end{cases} \quad \Rightarrow \quad \begin{cases} B' = B + \Delta B = B - B = 0 \\ F' = F + \Delta F = F - F = 0\,. \end{cases} \tag{8.61}$$

In a more compact form, we will simply write the gauge condition

$$B = 0 \qquad F = 0\,. \tag{8.62}$$

Notice that these two conditions determine ϵ^0 and ϵ^S completely, so small scalar gauge transformations are fully fixed by these requirements. The scalar part of the metric has then only diagonal perturbations, namely in h_{00} and h_{ii}. Traditionally these perturbations are called Φ and Ψ. So, with the identification $E = 2\Phi$ and $A = -2\Psi$, we find [87]

$$ds^2 = -\left(1 + 2\Phi\right) dt^2 + a^2 \left(1 - 2\Psi\right) dx^i \delta_{ij} dx^j\,. \tag{8.63}$$

This is the perturbed metric in Newtonian gauge [88] . Since in this gauge $B = F = 0$, the Einstein and Energy-momentum equations simplify considerably. Because of the SVT decomposition and are gauge choice, we can find the scalar equations by substituting

$$\begin{aligned} h_{00} &= -2\Phi\,, & \delta T_{00} &= 2\bar\rho\Phi + \delta\rho\,, \\ h_{0i} &= 0\,, & \delta T_{0i} &= -\left(\bar\rho + \bar p\right)\partial_i \delta u\,, \\ h_{ij} &= -a^2\delta_{ij}2\Psi\,, & \delta T_{ij} &= a^2\partial_{ij}\pi^S + \delta_{ij}a^2\left(\delta p - \bar p 2\Psi\right)\,, \end{aligned} \tag{8.65}$$

into the linearised Einstein equations. In particular, we have four equations corresponding to the trace of (8.9) (contracting with δ^{ij}), the traceless part of (8.9), (8.10) and (8.11), which

[87] Be aware that this is possibly the least universal convention in physics. You might find references where the definitions of Φ and Ψ as well as their signs are exchanged. Here we follow Weinberg's notation, which differ from Dodelson's notation by $\Phi_W = \Psi_D$ and $\Psi_W = -\Phi_D$.

[88] Be aware of the existence of the closely related *conformal* Newtonian gauge, defined such that

$$ds^2 = -a^2\left[\left(1 + 2\Phi\right) d\tau^2 + \left(1 - 2\Psi\right) dx^i \delta_{ij} dx^j\right]\,. \tag{8.64}$$

take the form

$$-\frac{1}{2M_{\rm Pl}^2}\left[\delta\rho-\delta p-\nabla^2\pi^S\right]=H\dot\Phi+\left(4H^2+2\frac{\ddot a}{a}\right)\Phi-\frac{\nabla^2\Psi}{a^2}+\ddot\Psi+6H\dot\Psi\,, \tag{8.66}$$

$$-\frac{a^2}{M_{\rm Pl}^2}\partial_i\partial_j\pi^S=\partial_i\partial_j\left(\Phi-\Psi\right)\,, \tag{8.67}$$

$$\frac{1}{2M_{\rm Pl}^2}\left(\bar\rho+\bar p\right)\partial_i\delta u=-H\partial_i\Phi-\partial_i\dot\Psi\,, \tag{8.68}$$

$$\frac{1}{2M_{\rm Pl}^2}\left(\delta\rho+3\delta p+\nabla^2\pi^S\right)=\frac{\nabla^2\Phi}{a^2}+3H\dot\Phi+3\ddot\Psi+6H\dot\Psi+6\frac{\ddot a}{a}\Phi\,. \tag{8.69}$$

The two scalar energy-momentum conservation equations ($T^{0\mu}_{;\mu}=0$ and the longitudinal part of $T^{i\mu}_{;\mu}=0$) are similarly obtained (see P.8.10)

$$\delta\dot\rho+3H\left(\delta\rho+\delta p\right)+\nabla^2\left[\frac{(\bar\rho+\bar p)}{a^2}\delta u+H\pi^S\right]-3\left(\bar\rho+\bar p\right)\dot\Psi=0\,, \tag{8.70}$$

$$\delta p+\nabla^2\pi^S+\partial_0\left[\left(\bar\rho+\bar p\right)\delta u\right]+3H\left(\bar\rho+\bar p\right)\delta u+\left(\bar\rho+\bar p\right)\Phi=0\,. \tag{8.71}$$

Note that (8.68) gives us δu simply as

$$\delta u=-\frac{H\Phi+\dot\Psi}{4\pi G_N(\bar\rho+\bar p)}\,. \tag{8.72}$$

A few comments are in order. First, notice also that although the energy-momentum conservation equations are not independent from the EE's, they contain one less derivative and therefore they are often more convenient to use. Second, the scalar constraint equation in this gauge is manifest in (8.67), which contains no time derivatives. In the absence of anisotropic stresses, a good approximation to our real universe, this equation is solved[89] by

$$\Phi=\Psi\quad\text{(no anisotropic inertia)}\,. \tag{8.73}$$

From the discussion around (A.35), we know that Φ corresponds to the non-relativistic Newtonian potential appearing in Newton's law of motion $\ddot x^i=-\partial^i\Phi$. Because of these two facts, both Ψ and Φ are often called *Newtonian* potentials.
Two combinations of the Einstein equations are particularly useful. The first is obtained by subtracting $(3/a^2)$ times (8.66) from (8.69) and using (8.67) to eliminate π^S and Φ. This gives a constraint on the valid initial conditions of the system,

$$a^3\,\delta\rho-3Ha^3(\bar\rho+\bar p)\,\delta u-\frac{a}{4\pi G}\nabla^2\Psi=0\,. \tag{8.74}$$

Because this equation does not contain time derivatives, it can be though of as a relativistic generalization of the non-relativistic Poisson equation, in Newtonian gauge. The second useful equation is obtained

[89]Notice that it is crucial to demand that Φ and Φ vanish at infinity for this solution to be unique.

Synchronous gauge* An alternative choice of gauge makes the temporal scalar part of the metric $h_{0\mu}$ vanish identically, namely one chooses ϵ^0 and ϵ^S such that

$$E = 0 \quad \text{and} \quad F = 0\,. \tag{8.75}$$

The perturbed metric takes the form

$$ds^2 = -dt^2 + a^2 dx^i dx^j \left[\delta_{ij}(1+A) + \partial_i\partial_j B\right]\,. \tag{8.76}$$

The clocks of observers at rest in these coordinates tick at the same rate, hence the name "synchronous". The four scalar Einstein equations are

$$-\frac{1}{M_{\rm Pl}^2}\left[\delta\rho - \delta p - \nabla^2\pi^S\right] = \frac{\nabla^2 A}{a^2} - \ddot{A} - 6H\dot{A} - H\nabla^2\dot{B} \tag{8.77}$$

$$-\frac{2}{M_{\rm Pl}^2}\partial_i\partial_j\pi^S = \partial_i\partial_j\left[\frac{A}{a^2} - \ddot{B} - 3H\dot{B}\right]\,, \tag{8.78}$$

$$\frac{1}{M_{\rm Pl}^2}\left(\bar\rho + \bar p\right)\partial_i\delta u = \partial_i\dot{A}\,, \tag{8.79}$$

$$-\frac{1}{M_{\rm Pl}^2}\left(\delta\rho + 3\delta p + \nabla^2\pi^S\right) = 3\ddot{A} + 6H\dot{A} + \nabla^2\ddot{B} + 2H\nabla^2\dot{B}\,. \tag{8.80}$$

The two scalar energy-momentum conservation equations ($T^{0\mu}_{;\mu} = 0$ and the longitudinal part of $T^{i\mu}_{;\mu} = 0$) are

$$\delta p + \nabla^2\pi^S + \partial_0\left[\left(\bar\rho + \bar p\right)\delta u\right] + 3H\left(\bar\rho + \bar p\right)\delta u = 0\,, \tag{8.81}$$

$$\delta\dot\rho + 3H\left(\delta\rho + \delta p\right) + \nabla^2\left[\frac{\left(\bar\rho + \bar p\right)}{a^2}\delta u + H\pi^S\right] + \frac{1}{2}\left(\bar\rho + \bar p\right)\partial_0\left[3A + \nabla^2 B\right] = 0\,. \tag{8.82}$$

Unlike for Newtonian gauge, the synchronous gauge conditions $E = 0 = F$ do not fix completely small gauge transformations. One can still perform a gauge transformation with

$$\epsilon_0 = -T(\mathbf{x}) \qquad \epsilon^S = a^2 T(\mathbf{x})\int\frac{dt'}{a(t')}\,, \tag{8.83}$$

which does not alter the condition $E = 0 = F$, but changes perturbations according to

$$\Delta\Psi = -\frac{\nabla^2 T}{a^2} - 3T\dot{H}\,, \qquad \delta u = T\,, \tag{8.84}$$

$$\Delta\delta\rho = -T\dot{\bar\rho} \qquad \Delta\delta p = -T\dot{\bar p}\,. \tag{8.85}$$

This additional redundancy can be fixed if the universe contains a non-relativistic fluid, such as for example dark matter. In that case, (8.81) tells us that δu_D is constant in time (up to corrections of order $\bar p_D/\bar\rho_D \ll 1$) and one can impose the additional gauge condition $\delta u_D = 0$, which completely fixes the gauge. To transform from synchronous to Newtonian gauge we can use (see P.8.11)

$$\Phi = -\frac{1}{2}\partial_0\left(a^2 B\right)\,, \qquad \Psi = -\frac{1}{2}A + \frac{a^2 H}{2}\dot{B}\,, \tag{8.86}$$

while the opposite conversion is W 5.3.46. A classic and extensive discussion of cosmological perturbation theory in Newtonian and syncronous gauges can be found in [41].

Comoving orthogonal gauge* Another option, often employed in the study of perturbation during inflation is comoving gauge[90], in which

$$\delta u = 0 \quad \text{and} \quad F = 0 \,. \tag{8.87}$$

It is straightforward to check that $\delta u = 0$ fixes ϵ^0, while ϵ^S is completely fixed by the condition $F = 0$. From its definition, the linearly perturbed energy momentum tensor is (W 5.1.43)

$$\delta T^i_j = \delta_{ij}\delta p + \partial_{ij}\pi^S + \partial_i \pi^V_j + \partial_j \pi^V_i + \pi^T_{ij} \,, \tag{8.88}$$

$$T^i_0 = \delta T^i_0 = \frac{\bar\rho + \bar p}{a^2}\left(a\partial_i F + aG_i - \partial_i \delta u - \delta u^V_i\right) \,, \tag{8.89}$$

$$T^0_i = \delta T^0_i = (\bar\rho + \bar p)\left(\partial_i \delta u + \delta u^V_i\right) \,. \tag{8.90}$$

Neglecting vector modes, $G_i = 0 = \delta u^V_i$, we find that in this gauge $T^i_0 = T^0_i = 0$. The fact that $T^i_0 = 0$ means that observers at rest in these coordinates are comoving with the fluid, while the fact that $T^0_i = 0$ means that the velocity of the fluid is orthogonal to the constant time hypersurfaces. Notice that in general in this gauge $\delta\rho \neq 0$.

Constant density gauge* This is another useful, but less used gauge for inflationary perturbations. As the name suggests, one imposes

$$\delta\rho = 0 \quad \text{and} \quad F = 0 \,. \tag{8.91}$$

These conditions fix the small gauge completely.

Spatially flat gauge* One last option we want to mention is to fix the spatial part of the metric to be completely unperturbed, $g_{ij} = a^2\delta_{ij}$, so that $h_{ij} = 0$. In the SVT notation one imposes

$$A = 0 \quad \text{and} \quad B = 0 \,. \tag{8.92}$$

In this gauge of course $E, F \neq 0$. But these (and more generally all $h_{0\mu}$) are non-dynamical degrees of freedom, since they appear with at most first derivatives in the EE's and the initial condition $\dot h_{0\mu}$ cannot be specified arbitrary but is fixed by the other initial conditions. So in some sense all dynamical scalar degrees of freedom in this gauge are in the matter sector as opposed to the metric sector.

8.6 Adiabatic modes

As the reader might have painfully noticed, even at linear order and for a single fluid, the equations of motion are already quite lengthy. Things get much worse when one includes all relevant constituents of the universe and/or goes beyond linear order. In practice this is often done with numerical codes such as CLASS or CAMB (one of the first efficient and popular code was CMBFAST). These codes, often collectively referred to as *Boltzmann codes*, are routinely used in data analysis and theoretical forecasting. To solve the equations of motion one also needs initial condition. In the currently favored cosmological model, initial conditions are set

[90] Notice that here again there is some confusion in the literature for the use of the term comoving.

up during a phase of very fast accelerated expansion in the first fraction of a second (inflation), as we mentioned in Sec. 6. One problem immediately arises when we try to evolve these initial condition forward in time since we do not know the constituents of the universe at energies much bigger than those probed at colliders, say above 10 TeV. Luckily for us, there seems to be quantities that, under certain conditions, are conserved and therefore can be trivially evolve in time. This result, which we are about to discuss, is one of the most important in cosmology. It allows one to study high energy physics by looking at the distribution of galaxies or of sub-eV photons. This remarkable connection of low-energy observables to high-energy physics has been a tremendous drive for the field of cosmology and has open new possibility to explore the fundamental laws of nature.
Let us start by introducing two new variables[91]

$$\mathcal{R} \equiv \frac{A}{2} + H\delta u\,, \qquad\qquad \zeta \equiv \frac{A}{2} - H\frac{\delta\rho}{\dot{\bar{\rho}}}\,. \tag{8.93}$$

From the gauge transformations, it is straightforward to check that both $\mathcal{R}$ and ζ are gauge invariant at linear order. We will refer to $\mathcal{R}$ as *curvature perturbations on comoving hypersurfaces*, because in comoving gauge $\mathcal{R} = A/2$ and A modifies the spatial diagonal part of the metric. For the same reason, ζ is often called *curvature perturbations on constant density hypersurfaces*. The two gauge-invariant variables are related at linear order due to the equations of motion. This is most easily seen in Newtonian gauge

$$\zeta(\vec{k},t) = \mathcal{R}(\vec{k},t) + \frac{M_{\rm Pl}^2}{3a^2(\bar{\rho}+\bar{p})}k^2 A(\vec{k},t) \quad \text{(Newtonian gauge)}\,. \tag{8.94}$$

Notice that the difference $\zeta - \mathcal{R}$ is proportional to $(k/aH)^2$, and therefore is negligible outside the Hubble radius, namely for $k_{\rm phy} = k/a \ll H$. So $\mathcal{R}$ will be conserved outside the Hubble radius if ζ is, and viceversa.

Adiabatic mode theorem We are now ready to state an important theorem [65]:

Whatever the constituents of the universe and outside the Hubble radius, $k \ll aH$, there are two conserved scalar adiabatic modes, i.e. $\dot{\mathcal{R}} = 0$, one of which satisfies $\mathcal{R} \neq 0$, and one conserved tensor mode, i.e. $\dot{D}_{ij} = 0$, for which $D_{ij} \neq 0$.

This statement is valid to all orders in perturbation theory around a flat FLRW spacetime, but we will prove it only at linear order. Also, we will work in Newtonian gauge[92]. Consider the following large gauge transformation that maintains Newtonian gauge (see P.8.12)

$$\epsilon_\mu = \left\{\epsilon(t), a^2\omega_{ij}x^j\right\}\,, \tag{8.95}$$

with ϵ some time dependent by space independent function and ω_{ij} an arbitrary spacetime 3×3 constant matrix. Since ϵ^μ does not vanish at spatial infinity, its existence does not contradict

[91] Notice that, unfortunately, different conventions for the names of these variables exists. A useful summary of the many possible choices in the literature is given in App A of [63].

[92] The theorem of can be proven in other gauges as well. In the original paper [65], Newtonian and synchronous gauges are discussed. In [20] and [35] the same derivation was presented for comoving gauge (aka "ζ-gauge") and generalized to higher order in derivatives.

the statement that Newtonian gauge conditions completely fixes the small gauge. If we start from an unperturbed flat FLRW universe, after this gauge transformation we find

$$\begin{aligned} &\Phi = -\dot{\epsilon}\,, \quad \Psi = H\epsilon - \frac{1}{3}\omega_{ii}\,, \\ &\delta p = -\dot{\bar{p}}\epsilon\,, \quad \delta\rho = -\dot{\bar{\rho}}\epsilon\,, \quad \delta u = \epsilon\,, \quad \pi^S = 0\,, \\ &D_{ij} = -\omega_{(ij)} + \frac{2}{3}\delta_{ij}\omega_{kk}\,. \end{aligned} \tag{8.96}$$

Notice that these transformation are completely different from those valid for small gauge transformations, Eq. (8.42), for example, the tensor perturbations D_{ij} is not invariant and so on. Notice that the anti-symmetric part of ω_{ij} is irrelevant since it does not generate any perturbation. Now comes the first crucial point. Since GR is a diff invariant theory and we started from unperturbed FLRW plus unperturbed $\bar{T}_{\mu\nu}$, which is a solution, the perturbations in Eq. (8.96) must be a solution of the equations of motion. This is also easily verified (see P.8.13). Recall that ϵ and ω do not vanish at spatial infinity, so this solution is an unphysical one. After all, it is just a change of coordinates.

The clever insight of Weinberg is to demand whether this gauge transformation can be extended to a physical solution. This is most easily though about in Fourier space, where the perturbations in Eq. (8.96) are all proportional do $\delta_D(\vec{k})$ and its derivative. A physical solution must eventually vanish at infinity and so its Fourier transform must be supported at $\vec{k} \neq 0$. When $\vec{k} \neq$ we are not guaranteed anymore that Eq. (8.96) is a solution. For all equations of motion that do not vanish as $\vec{\to} 0$, we know that a small modification of Eq. (8.96) is still a solution. For example, for the tensor perturbations, one can look for a solution of the form $D_{ij}(t) + \delta D_{ij}(t, \vec{k})$, where $D_{ij}(t)$ is the large perturbation in Eq. (8.96), and $\delta D_{ij}(t, \vec{k})$ is a small spatially varying (supported at $\vec{k} \neq 0$) correction. Given that we are solving linear differential equations, we can always find one such δD_{ij}. So we conclude that, whatever the constituents of the universe, there is always a solution to the equations of motion with a constant, non-vanishing D_{ij}, up corrections suppressed by k^2 in the superHubble limit. This solution represent the conservation of *primordial gravitational waves*. As we will discuss with inflation, the existence of this solution constitutes a unique opportunity to probe GR and its perturbative quantization.

The extension to a physical, non-constant solution can therefore be obstructed only when a given equation of motion vanishes identically for $\vec{k} = 0$. This happens for the off-diagonal part of the space-space Einstein equations, Eq. (8.67). We need therefore to check that this equation is solved also for $\vec{k} \neq 0$, namely

$$k_i k_j\,(\Phi - \Psi) = 0 \quad \Rightarrow \quad \Phi = \Psi\,. \tag{8.97}$$

This physicality condition fixes ϵ in terms of ω_{kk} as

$$\dot{\epsilon} + H\epsilon = \frac{1}{3}\omega_{kk} \quad \Rightarrow \quad \epsilon(t) = \frac{\omega_{kk}}{3a(t)} \int_T^t a(t')dt'\,, \tag{8.98}$$

where T represents some integration constant. Using this solution for ϵ and the perturbations in Eq. (8.96), we find

$$\mathcal{R} = \frac{\omega_{kk}}{3}\,. \tag{8.99}$$

We conclude that a solution with $\dot{\mathcal{R}} = 0$ and $\mathcal{R} \neq 0$ must always exist as consequence of diffeomorphism invariance. In other words, there is always a physical solution with constant $\mathcal{R}$ that sits nearby a gauge transformation. Before concluding, notice that since integration constant T is arbitrary, there is actual a second solution given by the different of two solutions with different T. This solution is

$$\Phi = \Psi = \frac{CH(t)}{a(t)}, \qquad \frac{\delta s}{\dot{\bar{s}}} = -\delta u = -\frac{\mathcal{R}}{a}, . \tag{8.100}$$

It decays with time during the hot big bang.

Adiabatic modes Notice that this procedure gives us the solution for metric perturbations

$$\Phi = \Psi = \mathcal{R}\left[-1 + \frac{H}{a}\int_T^t a(t')dt'\right], \tag{8.101}$$

and for fluid perturbations

$$\frac{\delta s}{\dot{\bar{s}}} = -\delta u = -\frac{\mathcal{R}}{a}\int_T^t a(t')dt', \tag{8.102}$$

for any diff scalar s (such as ρ and p). While it is already remarkable that we were able to find such a general solution, it is completely astounding that this solution appears to describe our universe Out of the many possible solutions one can find for the many fluids and metric perturbations, this adiabatic mode happens to be the one chosen by our universe. So lutions other than the adiabatic mode are often called isocurvature perturbations and have been searched for in the data but none has been found so far. Current bounds says that deviations from the adiabatic mode initial condition cannot be larger than about 1%, with small variations depending on the particular nature of each isocurvature mode. To drive this point home, for a given constituent a with density perturbation $\delta\rho_a$ we have

$$\frac{\delta\rho}{\dot{\bar{\rho}}} = -\frac{\delta\rho_a}{3H(\bar{\rho}_a + \bar{p}_a)} = -\frac{\delta\rho_a}{\bar{\rho}_a}\frac{1}{3H(1+w_a)}, \tag{8.103}$$

where we used the continuity equation for $\bar{\rho}$ and then assumed a is barotropic with equation of state parameter w_a. This means that for all constituents of the Universe at all times irrespectively of the time evolution on super-Hubble scales in our universe we have

$$\frac{\delta\rho_a}{(\bar{\rho}_a + \bar{p}_a)} = \frac{\delta\rho_b}{(\bar{\rho}_b + \bar{p}_b)}. \tag{8.104}$$

For a single component universes we can say a bit more. Assuming the Universe is dominate by a substance with constant equation of state parameter $w_t = p/\rho$, where "t" stands for "total", we can solve for the time dependence of the scale factor, $a = (t/t_0)^{2/(3w_t+1)}$ and perform the time integral in (8.101)

$$\frac{H}{a}\int_T^t a(t')dt' = \frac{2}{5+3w_t}. \tag{8.105}$$

This in turn gives

$$\Phi = \Psi = -\mathcal{R}\frac{3\,(1+w_t)}{5+3w_t}\,. \tag{8.106}$$

$$H\frac{\delta s}{\dot{s}} = -H\delta u = -\mathcal{R}\frac{2}{5+3w_t} = \Phi\frac{2}{3(1+w_t)}\,. \tag{8.107}$$

for single fluid backgrounds. While (8.101) is always valid, these expressions are (approximately) valid only when the universe is (approximately) dominated by a single component. These adiabatic modes will be the initial conditions we will use to study the formation of Large Scale Structures and the Cosmic Microwave Background. Finally, we will later see how quantum fluctuations during inflation generate precisely these modes.

8.7 Irreducible representations of ISO(3)*

The following discussion below paraphrases [67] Chapter 2, and we could not find an equivalent discussion in the literature). To find the irreps of ISO(3) we need to find a set of matrices $U(R,\alpha)$ for each ISO(3) element $\{R^i_j, \alpha_l\}$ that act on some Hilbert (vector) space of perturbations. In the following we will borrow the language from Quantum mechanics and refer to perturbations as "states" or "state-vectors". To begin, we note that "the component of the three-momentum all commute with each other and so it is natural to express physical state-vectors in terms of eigenvectors of the three-momentum." [67]. This is the usual Fourier transform: we consider state-vectors that are eigen-functions of translations

$$\hat{P}^i\psi_{k\sigma} = k^i\psi_{k\sigma}\,, \tag{8.108}$$

where σ is some other (discrete) quantum number that we have to figure out. Translations are represented by the unitary transformation

$$U(1,\alpha)\psi_{k\sigma} = e^{-ik^i\alpha_i}\psi_{k\sigma}\,. \tag{8.109}$$

Now, we want to find the action of rotations $U(R,0)\equiv U(R)$. Using the group properties, we note that

$$U(R)\psi_{k\sigma} = C_{\sigma\sigma'}(R,k)\psi_{Rk\sigma'}\,, \tag{8.110}$$

that is, a rotation changes the three-momentum of the state. We want now to find irreducible $C_{\sigma\sigma'}$ (i.e. that cannot be decomposed into smaller blocks by changing the basis for $\psi_{k\sigma}$). For this we will use the method of induced representations. The subgroup of ISO(3) we will be interested in is SO(3). The only invariant under SO(3) is the norm of a vector (and any function thereof), $k^ik^j\delta_{ij} = k^2$. Let us play some algebraic tricks now. For a reference vector q^i, define the rotation $S(k)$ that transforms it into any other vector k^i as

$$S(k)q = k \quad \Rightarrow \quad S^{-1}(k)k = q\,. \tag{8.111}$$

We can then re-write any state with momentum k as a transformation of a state with reference momentum q,

$$\psi_{k\sigma} = U(S(k))\psi_{q\sigma}\,. \tag{8.112}$$

Then, the action of a general rotation R can be massaged as follows:

$$\begin{aligned} U(R)\psi_{k\sigma} &= U(R)U(S(k))\psi_{q\sigma} && (8.113)\\ &= U(S(Rk))U(S^{-1}(Rk)RS(k))\psi_{q\sigma} && (8.114)\\ &= U(S(Rk))D_{\sigma\sigma'}\psi_{q\sigma'} && (8.115)\\ &= D_{\sigma\sigma'}U(S(Rk))\psi_{q\sigma'} && (8.116)\\ &= D_{\sigma\sigma'}\psi_{Rk\sigma'}\,, && (8.117)\\ & && (8.118) \end{aligned}$$

where in the third line we recognised that $S^{-1}(Rk)RS(k)q = q$ and so

$$U(S^{-1}(Rk)RS(k))\psi_{q\sigma} \equiv D_{\sigma\sigma'}\psi_{q\sigma}\,, \tag{8.119}$$

i.e. it must be some linear combination $D_{\sigma\sigma'}$ of states with momentum q. From this definition of $D_{\sigma\sigma'}$, we see that is it provides a representation of the little group, namely the subgroup of SO(3) that leaves the representative vector q invariant. For every little group rotation r, we have

$$U(r)\psi_{q\sigma} = D_{\sigma\sigma'}(r)\psi_{q\sigma'}\,. \tag{8.120}$$

Summarising, choosing a representative vector q and given a representation $D_{\sigma\sigma'}$ of the little group for q, we get a representation of the full group ISO(3) defined by

$$\begin{aligned} U(1,\alpha)\psi_{k\sigma} &= e^{-ik^i\alpha_i}\psi_{k\sigma}\,,\\ U(R,0)\psi_{k\sigma} &= D_{\sigma\sigma'}(r(R,k))\psi_{Rk\sigma'}\,, \end{aligned} \tag{8.121}$$

where the little group element $r(R,k)$ is given by

$$r(R,k) \equiv S^{-1}(Rk)RS(k)\,. \tag{8.122}$$

Little groups*

While for the Poincaré group there are 6 little groups, of which only three have physical significance (the vacuum, massive particles and massless particles), for cosmology there are only two little groups: SO(3) itself for $q^iq_i = 0$, and SO(2) for $q^iq_i \neq 0$.

The irreps of SO(3) are well known from the study of angular momentum in quantum mechanics. They are classified by the Casimir operator J^2, with eigen-values $l(l+1)$ for $l = 0, 1/2, 1, \ldots$ and are of dimension $2l+1$ with states $|l,m\rangle$ and $|m| \leq l$. Focussing on the bosonic irreps with integer l, we know they correspond to spin zero, one, two, etc. The field operators that generate those states are:

$$\begin{aligned} &\text{Spin zero:} && \phi,\ h_{ii},\ \ldots && (8.123)\\ &\text{Spin one} && h_{0i},\ u_i,\ \ldots && (8.124)\\ &\text{Spin two:} && h_{\langle ij\rangle} \equiv h_{ij} - \frac{1}{3}h_{kk}\delta_{ij},\ \ldots\,. && (8.125) \end{aligned}$$

Notice that the splitting between the trace of the two-tensor h_{ij}, which has spin zero, and its traceless part $h_{\langle ij\rangle}$, which has spin two, is purely algebraic and does not involve any (inverse)

Laplacians. These $q = 0$ irreps are relevant to classify and discuss the background and adiabatic modes. For physical perturbations, we have to consider the other representative vector.
For $q^i q_i \neq 0$, we can choose as representative vector $q^i = \{q, 0, 0\}$ so that the little group is recognised as two-dimensional rotations, namely SO(2), which is an abelian group. All complex representations of an Abelian group are one-dimensional by Schur's lemma (all real representations are two dimensional). There are infinitely many such representations, enumerated by an integer $m \in \mathbb{N}$. Physically, we can interpret m as the "helicity" of the state, i.e. how it transforms under a rotation around the direction of its momentum. If the underlying theory is parity invariant, which is sometimes assumed in cosmological applications, for every state with helicity m there as to exist a state of helicity $-m$. So we have classify states as helicity 0, 1, 2 etc.

Problem lesson 8

P.8.1 *Optional* Find a computer with Mathematica. Install xAct and xPand following the instructions here, and use it to derive the linearised Einstein Equations in any gauge, as given in the notes.

P.8.2 Solve Eq. (8.25) for $\theta(v)$ (not $\theta(v, \omega)$). From the solution, assuming that v_i transforms as tensors under diffeomorphism (and therefore also under rotations), show explicitly that θ transforms as a scalar under rotations $\theta'(x', t) = \theta(x, t)$. Does θ transform as a scalar also under general diffs?

P.8.3 Extract all the 4 tensors, 2 transverse vectors and the transverse traceless two-tensor from the a generic symmetric two-tensor $T_{\mu\nu}$. It is sufficient to write down an appropriate number of differential equations satisfied by these objects, you do not need to write the solutions of those equations (which is anyways straightforward). To achieve this, you might want to consider acting on the tensor with various combinations of one and two spatial derivatives ∂_i.

P.8.4 Derive the gauge transformation for vectors and two-tensors Eq. (8.38) and Eq. (8.39), at linear order in ϵ^μ.

P.8.5 A change of coordinates $x'^\mu = x^\mu + \epsilon^\mu(x)$ induces a *gauge transformation* on all perturbations. In particular, the vector perturbations in the metric C_i and G_i, defined in Eq. (8.27), transform according to Eq. (8.42). Find a combination of C_i and G_i that is invariant under gauge transformations. It will help to think about the mass dimension of these two perturbations. Compare the gauge invariant combination with the equations for vectors (8.47) (see also W 5.1.50-52).

P.8.6 Compute the equation of motion for a massless scalar field, with action

$$S = \int d^4x \sqrt{-g}\, \frac{1}{2} \partial_\mu \phi \partial^\mu \phi \,. \tag{8.126}$$

Compare it with the equation of motion for the tensor modes in Eq. (8.49), aka gravitational waves.

P.8.7 Assuming $\pi_{ij}^T = 0$, solve the tensor equations of motion well inside and well outside the Hubble radius, $k \gg aH$ and $k \ll aH$ respectively.

P.8.8 *Optional* Compute the gauge transformations of the components of the metric A, B, C_i, D_{ij}, E, F and G_i and the analogous SVT components of the energy momentum tensor. You should reproduce Eq. (8.42) and Eq. (8.43). What do you need to assume about the scaling of $\epsilon^\mu(x)$ for $\mathbf{x} \to \infty$?

P.8.9 *Optional* Verify that the actual eom's Eq. (8.45), Eq. (8.47) and Eq. (8.49) are indeed of the form Eq. (8.18). Perform a Fourier transform and check that indeed different Fourier modes decouple.

P.8.10 Derive the continuity equation in Newtonian gauge, Eq. (8.70)

P.8.11 Derive the conversion formulae from synchronous to Newtonian gauge. You should reproduce W 5.3.51-52.

P.8.12 Prove that the transformation Eq. (8.95) maintains the Newtonian gauge conditions, namely the form of the metric in Eq. (8.63). Beware that since Eq. (8.95) represents a *large* gauge transformation (it does not vanish at spatial infinity), one can still use the general gauge transformations Eq. (8.39) but *not* those in Eq. (8.42), which had been derived only for small gauge transformations, which vanish at infinity.

P.8.13 Verify that Eq. (8.96) are solutions of the Newtonian gauge equations of motion.

9 Correlators and initial conditions

9.1 Introduction to the observable fields

Large scale structures Over the past decades, galaxy redshift surveys have mapped an increasingly large fraction of the observable universe, revealing that galaxies are not uniformly distributed but form a rich web of filaments, sheets and voids. This so–called large–scale structure (LSS) arises from the small primordial fluctuations whose statistical properties will be described in following sections. When averaged on sufficiently large scales, e.g. hundreds of Mpc, the Universe again appears statistically homogeneous and isotropic, in agreement with the FLRW background.
To quantify the deviations from exact homogenity, the distribution of all constituents can be described in terms of the fractional overdensity field δ. For example, for the density of matter, which includes baryons and dark matter, we define

$$\delta_m(\vec{x},t) \equiv \frac{\rho_m(\vec{x},t) - \bar{\rho}_m(t)}{\bar{\rho}_m(t)}\,, \tag{9.1}$$

where $\rho_m(\vec{x},t)$ is the local matter density and $\bar{\rho}_m(t)$ its spatial average. Because of the homogeneity of the background, we will work most of the time in Fourier space, where different Fourier mode decouple from each other at linear order

$$\delta_m(\mathbf{k},t) = \int d^3x\, e^{-i\mathbf{k}\cdot\mathbf{x}}\delta_m(\mathbf{x},t)\,. \tag{9.2}$$

Since the matter density is dominated by dark matter, which interacts only gravitationally, it is hard to observed directly. What we actually observe are the bright tracers of the matter field such as galaxies, including quasars. Other less bright tracers can also be seen such as neutral hydrogen or carbon monoxide. For galaxies, the spatial distribution can be described by the dimensionless perturbation to their comoving number density,

$$\delta_g(\vec{x},t) \equiv \frac{n_g(\vec{x},t) - \bar{n}_g(t)}{\bar{n}_g(t)}\,, \tag{9.3}$$

where $n_g(\vec{x}), t$ denotes the local number density of galaxies and $\bar{n}_g(t)$ its mean value.
On sufficiently large scales, where gravitational evolution remains in the linear regime and the formation of galaxies is effectively local, the galaxy overdensity δ_g is empirically found to be proportional to the underlying matter overdensity:

$$\delta_g(\vec{x},t) = b(t)\,\delta_m(\vec{x},t)\,, \tag{9.4}$$

where $b(t)$ is the possibly time-dependent *bias parameter.* This simple approximate relation is called *linear, local bias.* In general, galaxy bias is both non–linear and non–local, reflecting the complex physics of galaxy formation and feedback. However, on very large scales these corrections are expected to be small, and in these notes we will restrict to this simplest approximate, which provides a good balance of simplicity and accuracy.
The statistical properties of δ_m can then be computed within linear perturbation theory, and related to observables t such as the galaxy two–point correlation function or power spectrum, as we discuss in the following.

The cosmic microwave background A second, and even more direct, probe of primordial perturbations is provided by the cosmic microwave background (CMB). About 4×10^5 years after the Big Bang, the universe cooled sufficiently for protons and electrons to combine into neutral hydrogen, allowing photons to travel freely for the first time. These photons constitute Tiny anisotropies in its temperature and polarisation encode a snapshot of the universe at the time of recombination.
A generic electromagnetic signal should be described in terms of its intensity $I_\nu(\hat{n})$ in the direction $\hat{n}$ in the sky at frequency ν. Since the CMB forms a nearly perfect blackbody with temperature $T_0 \simeq 2.73$ K the frequency dependence is fixed and we simply focus on the direction dependence. Therefore, we describe the observed CMB as a temperature field as a function of direction $\hat{n}$ on the sky,

$$T(\hat{n}) = \bar{T}_0 \left[1 + \Theta(\hat{n})\right], \tag{9.5}$$

where $\Theta(\hat{n})$ denotes the fractional temperature fluctuation. The temperature field $\Theta(\hat{n})$ is defined on the two–dimensional sphere and can therefore be decomposed in the complete orthonormal basis of spherical harmonics $Y_{\ell m}(\hat{n})$,

$$\Theta(\hat{n}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m}\, Y_{\ell m}(\hat{n}) \,. \tag{9.6}$$

The coefficients $a_{\ell m}$ quantify the amplitude of fluctuations on angular scales of order π/ℓ and encode all information about the angular structure of the CMB sky. Large–scale features, corresponding to small ℓ, describe fluctuations on tens of degrees, while small–scale anisotropies correspond to large ℓ values, down to the arcminute range or less.
The spherical harmonic coefficients can be obtained from the observed temperature map by the inverse transform,

$$a_{\ell m} = \int d\Omega_{\hat{n}}\, Y^*_{\ell m}(\hat{n})\, \Theta(\hat{n}) \,, \tag{9.7}$$

where $d\Omega_{\hat{n}}$ is the element of solid angle. This decomposition plays a role analogous to the Fourier transform in flat space: it expresses the temperature field as a sum of orthogonal angular modes.
On sufficiently small patches of the sky, where curvature of the celestial sphere can be neglected, one can use the *flat–sky approximation.* In this limit, the direction on the sky is described by a two–dimensional vector $\boldsymbol{\theta}$ and the temperature field can be expanded as a Fourier transform,

$$\Theta(\boldsymbol{\theta}) = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2}\, \tilde{\Theta}(\boldsymbol{\ell})\, e^{i\boldsymbol{\ell}\cdot\boldsymbol{\theta}} \,, \tag{9.8}$$

with inverse relation

$$\tilde{\Theta}(\boldsymbol{\ell}) = \int d^2\boldsymbol{\theta}\, \Theta(\boldsymbol{\theta})\, e^{-i\boldsymbol{\ell}\cdot\boldsymbol{\theta}} \,. \tag{9.9}$$

Here $\boldsymbol{\ell}$ is the two–dimensional analogue of the multipole vector, and its magnitude $\ell = |\boldsymbol{\ell}|$ corresponds to the angular scale $\theta \sim \pi/\ell$. The flat–sky approximation is extremely accurate for $\ell \gtrsim 50$, and it allows many CMB analyses to be performed using standard Fourier methods familiar from studies of large–scale structure.

9.2 Statistical initial conditions

What we are able to measure with cosmological observations is the distribution of stuff in the universe. More precisely, if δ represents the density of matter, galaxy of of CMB photons, or even more generically perturbations to the metric and of the energy-momentum tensor of all other constituents of the universe, then cosmological observations gives us partial information on $\delta(\mathbf{x}, t)$ on our past light cone. Unfortunately, no theory can predict this quantity. There are a few reason for this. First, quantum mechanics, which as we will see underlies the generation of perturbation in our universe, gives statistical predictions and so can only give us probabilities for a give $\delta(\mathbf{x}, t)$. Second, all fundamental laws of nature and the FLRW background are invariant under translations, and so any expectation value $\langle \delta(\mathbf{x}, t) \rangle$ we compute would not depend on $\mathbf{x}$. What we can instead predict are averages n-point correlation functions, in real or Fourier space

$$\langle \prod_{a=1}^{n} \delta(t_a, \mathbf{x}_a) \rangle = \int_{\mathbf{k_1}} \int_{\mathbf{k_2}} \dots \int_{\mathbf{k_n}} e^{i \sum_a^n \mathbf{k}_a \cdot \mathbf{x}_a} \langle \prod_{a=1}^{n} \delta(t_a, \mathbf{k}_a) \rangle \,. \tag{9.10}$$

At the most fundamental level, these averages $\langle \dots \rangle$ represent *quantum expectation values*[93] computed on the quantum state of universe $|\Omega\rangle$

$$\langle \Omega | \, \hat{\delta}(x_1) \hat{\delta}(x_2) \dots \hat{\delta}(x_n) \, | \Omega \rangle \ , \tag{9.11}$$

where the hats have been introduced to stress that all the field should be promoted to operators. In practice though, these correlator lose their quantum character as soon as the relevant perturbations become large than the Hubble radius, which happens during inflation. For the remaining of the cosmological history, quantum perturbations are expected to be very well[94] approximated by classical or stochastic correlators. This in particular implies that for all non-inflationary cosmological applications we can treat all fields as classical *commuting fields* whose averages are predicted by a probability distribution functional as opposed to a wavefunctional. This is the setup of a *stochastic* field theory, where the fields at each spacetime location x is a random variable. While in the next subsection we make some very general statements valid for all types of averages, for the rest of these notes we will restrict ourselves to the simplest possible distribution, that of a Gaussian random field, because this is still a perfect fit to all data captured by linear order perturbation theory.

Statistical homogeneity and isotropy The assumption that the fundamental theory of nature is Poincaré invariant, plus our identification of a homogeneous and isotropic FLRW background as starting point of perturbation theory allow us to make some general statement about the properties of the correlators in Eq. (9.10). Correlators have to be *statistically homogeneous and isotropic*, namely

$$\langle \prod_{a=1}^{n} \delta(t_a, R \cdot \mathbf{x}_a + \mathbf{a}) \rangle = \langle \prod_{a=1}^{n} \delta(t_a, \mathbf{x}_a) \rangle \tag{9.12}$$

[93] Notice that these are not scattering amplitudes, aka “in-out” correlators, but rather “in-in” correlators. They are related to the perhaps more familiar S-matrix formalism by the LSZ reduction formula

[94] We will discuss this quantitatively in Sec. **??**.

for every constant spatial vector $\mathbf{a}$ and constant rotation $R \cdot x = R^i{}_j x^j$ in $SO(3)$. The infinitesimal form of this constraints is sometime useful. Consider the real space generators of spatial translation and rotations

$$P_i : -\partial_i \quad \text{and} \quad R_{ij} : -\left(x_i \partial_j - x_j \partial_i\right) . \tag{9.13}$$

Because the theory and the background respect homogeneity and isotropy, these generators must annihilate any correlator. More in detail, the such of the action of these generators on the arguments of each perturbation $\delta(\mathbf{x}, t)$ (assumed to be a scalar for simplicity) must vanish

$$\sum_{a=1}^{n} \frac{\partial}{\partial \mathbf{x}_a} \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\dots\delta(\mathbf{x}_n)\rangle \stackrel{!}{=} 0 \,, \tag{9.14}$$

$$\sum_{a=1}^{n} \left(x_a^i \frac{\partial}{\partial x_a^j} - x_a^j \frac{\partial}{\partial x_a^i} \right) \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\dots\delta(\mathbf{x}_n)\rangle \stackrel{!}{=} 0 \,. \tag{9.15}$$

The general solution of the first constraint is that the correlator only depends on $n-1$ variables, for example $\mathbf{x}_a - \mathbf{x}_1$ for $a = 2, \dots n$. The second is simply solved if the correlator depends only on $SO(3)$ invariant objects, namely contractions of all covariant spatial indices with δ_{ij}. These generators can also be specified for the Fourier space correlators

$$P_i : -k_i \quad \text{and} \quad R_{ij} : -\left(k_i \partial_j - k_j \partial_i\right) , \tag{9.16}$$

and therefore

$$\sum_{a=1}^{n} \mathbf{k}_a \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\dots\delta(\mathbf{k}_n)\rangle \stackrel{!}{=} 0 \,, \tag{9.17}$$

$$\sum_{a=1}^{n} \left(k_a^i \frac{\partial}{\partial k_a^j} - k_a^j \frac{\partial}{\partial k_a^i} \right) \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\dots\delta(\mathbf{k}_n)\rangle \stackrel{!}{=} 0 \,. \tag{9.18}$$

The first condition is satisfied if the correlator is proportional to Dirac delta function, so we introduce the conventions (see P.9.1)

$$\langle \delta(t_1, \mathbf{k}_1)\delta(t_2, \mathbf{k}_2)\dots\delta(t_n, \mathbf{k}_n)\rangle \equiv (2\pi)^3 \delta_D^{(3)} \left(\sum_{a=1}^{n} \mathbf{k}_a \right) \langle \delta(t_1, \mathbf{k}_1)\delta(t_2, \mathbf{k}_2)\dots\delta(t_n, \mathbf{k}_n)\rangle' \,. \tag{9.19}$$

The condition Eq. (9.18) requires that the correlator depends only on the rotational invariant contractions $\mathbf{k}_a \cdot \mathbf{k}_b$. For equal time correlators one often employs the notation

$$\langle \prod_a^n \delta(t, \mathbf{k}_a)\rangle \equiv (2\pi)^3 \delta_D^{(3)} \left(\sum_{a=1}^{n} \mathbf{k}_a \right) B_n(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_n; t) \,. \tag{9.20}$$

Ensemble averages and spatial averages Before continuing we have to address the elephant in the room. In cosmology we are interested in the statistical properties of random fields such as the matter overdensity $\delta(\vec{x})$ or the CMB temperature fluctuation $\Theta(\hat{n})$. In principle, these statistics are defined as ensemble averages over many possible realisations of the universe, each corresponding to a different set of initial conditions. However, we only observe a single universe so how can we every observationally test our predictions? To make progress, we assume that spatial averages within our universe are representative of the ensemble averages defined by the underlying probability distribution. This assumption, known as the ergodic hypothesis, is justified by the ergodic theorem. The main assumption of the theorem is statistical homogeneity and that the correlations of a field decay at distances much larger than a characteristic correlation scale r_c. In formulae this means that if we separate a group of fields in two and move one subgroup at a distance $y \gg r_c$ then the correlators factorize, namely

$$\langle \prod_{i=1}^{n} \delta(x_i) \prod_{j=n+1}^{N} \delta(x_j + y) \rangle \simeq \langle \prod_{i=1}^{n} \delta(x_i) \rangle \langle \prod_{j=n+1}^{N} \delta(x_j + y) \rangle + \dots \tag{9.21}$$

as $y \to \infty$. In this case, spatial averages over distanced much larger than r_c are effectively sampling independent realizations of the fields and therefore give an accurate description of the actual statistical average.

9.3 Gaussian random fields

While for these notes we will only concern ourselves with the simplest possible statistical properties, namely those of a *Gaussian random field*, it is worth recalling a few basic facts about random variables.

Random Variables and their Statistical Descriptors A *random variable* X is a quantity that can take different values with some probability distribution. Its statistical properties are entirely characterised by the *probability density function* (PDF) $P(X)$, which satisfies

$$P(X) \geq 0, \qquad \int_{-\infty}^{\infty} dX\, P(X) = 1. \tag{9.22}$$

The ensemble average or *expectation value* of any function $f(X)$ is defined as

$$\langle f(X) \rangle = \int_{-\infty}^{\infty} dX\, P(X)\, f(X). \tag{9.23}$$

In particular, the *nth moment* of X is

$$\mu_n \equiv \langle X^n \rangle = \int dX\, P(X)\, X^n, \tag{9.24}$$

Here we will always assume the distribution is centered around zero, i.e. $\langle X \rangle = 0$. This can always be achieved for any variable simply by subtracting its average, $X \to X - \langle X \rangle$. The second moment give the familiar mean and variance, $\mu_2 = \sigma^2$.
It is often convenient to encode all moments in the *moment generating function*

$$M(\lambda) \equiv \langle e^{\lambda X} \rangle = \int dX\, P(X)\, e^{\lambda X}. \tag{9.25}$$

The corresponding *cumulant generating function* is defined as the logarithm of $M(\lambda)$,

$$K(\lambda) \equiv \ln M(\lambda) = \ln \langle e^{\lambda X} \rangle, \tag{9.26}$$

and its derivatives at $\lambda = 0$ yield the *nth cumulant*:

$$\kappa_n \equiv \left. \frac{d^n K(\lambda)}{d\lambda^n} \right|_{\lambda=0} . \tag{9.27}$$

The computation of moments from cumulant is familiar from QFT where it goes under the name of Wick's theorem. Assuming $\langle X \rangle = 0$ the first few cumulants are

$$\kappa_1 = 0, \qquad \kappa_2 = \langle X^2 \rangle, \qquad \kappa_3 = \langle X^3 \rangle, \quad \kappa_4 = \langle X^4 \rangle - 3\langle X^2 \rangle^2.$$

Note that the second cumulant and moment coincide for variables with zero average. Cumulants provide a convenient characterisation of a distribution: they add linearly for independent random variables. A crucial property for us is that for a Gaussian distribution, all cumulants beyond the second vanish. This makes precise the idea that a Gaussian distribution is fully characterised by its two point function.

Gaussian distributions The relevant fields for us are defined on a fixed time hypersurface, but it is straightforward to include a time dependence if desired. Let us start with a Gaussian random variable X, also called a *normal variable.* The PDF is

$$P(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{X^2}{2\sigma^2} \right) , \tag{9.28}$$

where σ^2 is the variance and we assumed a vanishing mean $\langle X \rangle = 0$. Such a variable is completely characterised by its variance: all higher–order cumulants vanish, so its statistics are entirely determined by the second moment

$$\langle X^{2n+1} \rangle = 0 \, , \quad \langle X^{2n} \rangle = \sigma^{2n} (2n-1)!! \, . \tag{9.29}$$

The cumulants of the distribution, a.k.a. the *connected correlators* $\langle \ldots \rangle_c$ are simply

$$\langle X^m \rangle_c = \begin{cases} \sigma^2 & m = 2 \, , \\ 0 & \text{otherwise} \, . \end{cases} \tag{9.30}$$

A *Gaussian random field* $\delta(\vec{x})$ is the spatial generalisation of a Gaussian random variable. It is defined as a field whose joint probability distribution for points $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N\}$ is a multivariate Gaussian:

$$P(\delta_1, \ldots, \delta_N) = \frac{1}{\sqrt{(2\pi)^N \det C}} \exp\left[-\frac{1}{2} \, \delta_i \, (C^{-1})_{ij} \, \delta_j \right] , \tag{9.31}$$

where $\delta_i \equiv \delta(\vec{x}_i)$ and $C_{ij} = \langle \delta(\vec{x}_i) \, \delta(\vec{x}_j) \rangle$ is the covariance matrix. A Gaussian random field is completely characterised by its two–point correlation function, or equivalently by its power spectrum. If the field has zero mean as we will assume, $\langle \delta(\vec{x}) \rangle = 0$, then all its odd moments vanish and all higher–order connected correlators (cumulants beyond second order) are zero. It is then easier to simply work with the correlators, namely

$$\langle \delta(\mathbf{x}) \rangle = 0 \, , \quad \langle \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \rangle = \xi(\mathbf{x}_1, \mathbf{x}_2) \, , \quad \langle \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \ldots \delta(\mathbf{x}_n) \rangle_c = 0 \, , \tag{9.32}$$

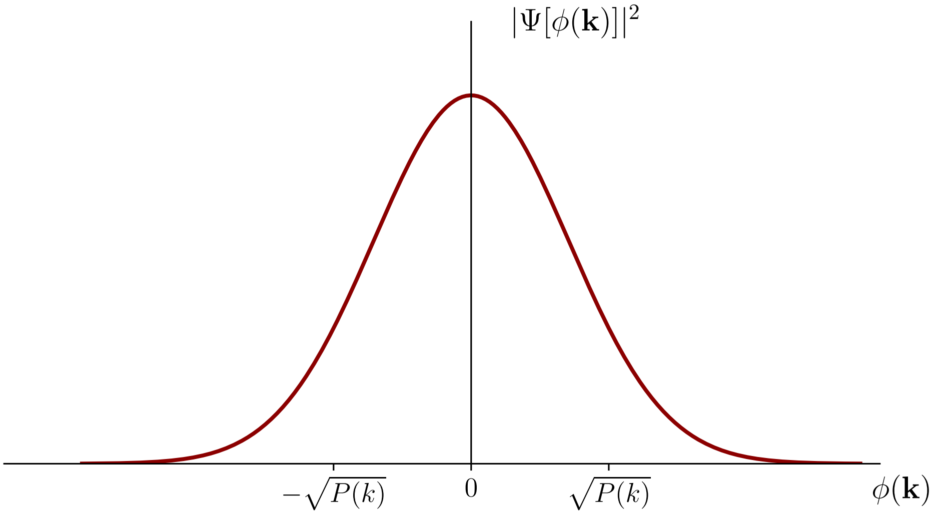


Figure 21: A Gaussian distribution.

for $n > 2$, where we defined the real space *correlation function* ξ. Imposing statistical homogeneity and isotropy, we find $\xi = \xi(|\mathbf{x}_1 - \mathbf{x}_2|)$. Since the Fourier transform is a linear operation, the same formulae apply mutatis mutandis to Fourier space correlators. Imposing again statistical homogeneity and isotropy, one finds

$$\langle\delta(\mathbf{k})\rangle = 0\,, \quad \langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\rangle = (2\pi)^3\delta_D^{(2)}\left(\mathbf{k}_1+\mathbf{k}_2\right)P(|\mathbf{k}_1|)\,, \quad \langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\ldots\delta(\mathbf{k}_n)\rangle_c = 0\,, \tag{9.33}$$

where we defined the *power spectrum* $P(k)$, which quantifies the typical size of density perturbations with wavenumber k, as depicted in Fig. 21. The power spectrum is simply related to the correlation function by

$$\xi(r) = \int_{\mathbf{k}}\int_{\mathbf{k}'}(2\pi)^3\delta_D^{(2)}\left(\mathbf{k}+\mathbf{k}'\right)P(|\mathbf{k}|)e^{i(\mathbf{k}\cdot\mathbf{x}+\mathbf{k}'\cdot(\mathbf{x}+\mathbf{r}))} \tag{9.34}$$

$$= \int\frac{dkk^2}{(2\pi)^3}d\cos\theta d\phi P(k)e^{ikr\cos\theta} = \int_0^\infty\frac{dkk^2}{2\pi^2}\frac{\sin(kr)}{kr}P(k)\,. \tag{9.35}$$

Of course one can take the power spectrum of different perturbations, such as $\mathcal{R}$ or δ, so sometimes we will specify it by writing $P_{\mathcal{R}}$ or P_δ. Notice that for a statistically homogeneous Gaussian field, every wavenumber $\mathbf{k}$ is independent from every other wavenumber,

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle = 0 \quad \text{for} \quad \mathbf{k}'\neq\mathbf{k}$$

It is only non-linear evolution, beyond linear perturbation theory, that generates correlation among different wavenumbers. The Fourier space three-point connected correlation function is often called *bispectrum* and the four-point function the *trispectrum*. they will not be discussed in these notes.

9.4 Initial conditions of the universe

Perturbation theory allows us to connect $\delta(t,\mathbf{k})$ at some time t to some initial condition $\delta(t_i,\mathbf{k})$. A remarkable fact about our Universe is that if we take t_i early enough that a given mode is superHubble, $k \ll aH$, then perturbations are in the adiabatic mode we derived in the last section, Eq. (8.101) and Eq. (8.102). In this mode, every perturbation is proportional to the primordial curvature perturbation $\mathcal{R}_p(\mathbf{k})$, which is time independent. So to linear order, every perturbation we have encountered so far can be written as

$$\delta(t,\mathbf{k}) = F(k,t)\mathcal{R}_p(\mathbf{k})\,, \tag{9.36}$$

for some transfer function $F(k,t)$ that depends only on the norm of the wavevector $k = |\mathbf{k}|$ and that is fully determined by solving the Einstein and continuity equations. In particular, all the statistical properties of all perturbations are fully determined by those of $\mathcal{R}(\mathbf{k})$. For example

$$\langle \prod_{a=1}^{n} \delta(\mathbf{k}_a, t_a) \rangle = \left[\prod_{b=1}^{n} F(k_b, t_b)\right] \langle \prod_{a=1}^{n} \mathcal{R}_p(\mathbf{k}_a)\rangle + \mathcal{O}(\mathcal{R}^2)\,. \tag{9.37}$$

Remarkably, we have observed that the statistics of the random variable $\mathcal{R}_p(\mathbf{x})$ is precisely the simplest possible one, namely that of a Gaussian random fields. This property has been tested to better than a part in ten thousand, depending of type of *non-Gaussianity* [2]. Finally, measurement of the anisotropies of the CMB and of the distribution of LSS have determined the function $P_{\mathcal{R}}(k)$ up to a few percent on cosmological scales $H_0 < k < 0.2$ Mpc^{-1} to be approximately scale invariant, namely

$$P_{\mathcal{R}}(k) \equiv \frac{\Delta_R^2(k) 2\pi^2}{k^3}\,, \quad \Delta^2_{\mathcal{R}}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s - 1}\,, \tag{9.38}$$

with [1]

$$\begin{aligned} A_s &= 2.2 \times 10^{-9} && \textit{primordial amplitude}\,, \\ n_s &= 0.965 \pm 0.005 && \textit{spectral tilt}\,, \\ k_* &= 0.05\ \text{Mpc}^{-1} && \textit{pivot scale}\,. \end{aligned} \tag{9.39}$$

Note that $P(k) \propto k^{-3-0.03}$ so the primordial power spectrum is almost scale invariant, an observation that led us to inflation before the hot big bang. Moreover, $\mathcal{R}$is a dimensionless quantity of typical size $\sqrt{A_s} \sim 10^{-5}$. This ensures that, as long as perturbations don't grow too much, i.e. $F \ll 10^5$, then linear perturbation theory should give excellent results. On scales of order Hubble or larger higher order corrections are indeed of order $\mathcal{R}^2 \sim 10^{-10}$ and hence highly suppressed.

In summary, the distribution of everything we observe at cosmological scales started with an initial condition proportional to $\mathcal{R}(\mathbf{x})$. On large scales this field appears to be a Gaussian random field. Therefore the statistical properties of all perturbations in the universe are determined by a single real function of a single variable, $P_{\mathcal{R}}(k)$, whose for is given above. In the next section we will learn how to precisely go from $\mathcal{R}$ to the matter perturbations δ_m and the CMB temperature anisotropies $T(\hat{n})$.

Problem lesson 9

P.9.1 Prove that all Fourier space correlators Eq. (9.10) are proportional to Dirac delta functions that enforce the conservation of (spatial) momentum by taking the Fourier transform of Eq. (9.12).

P.9.2 Derive Eq. (9.34)

P.9.3 Consider a Gaussian random field $\delta_G(\mathbf{x})$ and a non-Gaussian random field $\delta(\mathbf{x})$ related by

$$\delta(\mathbf{x}) = \delta_G(\mathbf{x}) + f_{NL}\delta_G(\mathbf{x})^2\,, \tag{9.40}$$

where f_{NL} is a constant that parameterize the deviation from Gaussianity (NL stands for "non-linear"). This particular type of non-Gaussianity if often called "local" non-Gaussianity, because all δ_G are evaluated at the same point $\mathbf{x}$. Compute the bispectrum of $\delta(\mathbf{k})$ to *linear* order in f_{NL}.

10 Newtonian Perturbation Theory and Structure Formation

The story of the universe is the tale of how perturbations started small and independent, i.e. Gaussian, and grew up to be large and correlated, i.e. non-Gaussian. Within linear perturbation theory, the story of this growth is conceptually simple and can be remembered using the following cartoonish equation of motion for matter density perturbations δ

$$\partial_t^2 \delta = [\text{gravitational attraction} - \text{pressure} - \text{expansion}]\, \delta \,, \tag{10.1}$$

Borrowing intuition from the standard harmonic oscillator, we see that pressure and the expansion of the universe tend to obstruct the growth of structures, while gravitational attraction tend to boost it. Depending on which of the three different forces dominate at every given time, perturbations evolve differently as summarized in Fig. **??**

- When pressure is strongest we recover the standard wave equation with oscillatory solutions, whose amplitude is approximately constant. This is the regime relevant at early time, which will play a crucial role in determining the structure of the anisotropies of the CMB.
- When cosmological expansion wins, perturbations can decay or remain approximately constant, depending on the rate of acceleration and deceleration. This regime is relevant for dark matter during radiation domination and dark energy domination.
- When gravitational attraction is strongest, gravitational collapse takes place. Unlike what one might naively expect in a non-expanding universe, or from the harmonic oscillator analogy above, this collapse is usually not exponential but rather power law in time on cosmological scales, because of the regulating action of the expansion of the universe. This is the regime relevant for the formation of structures and eventually determines the distribution of dark matter and galaxies.

As perturbations δ grow (or as we require more precise predictions), the linear approximation becomes insufficient. For a while this can be fixed by including higher and higher orders in perturbation theory and this is an active research field. As δ becomes of order one, perturbation theory breaks down and we need to resort to numerical solutions of the equations of motion or, more commonly, to N-body simulations. In this introductory course, we will limit ourselves to the linear regime, with only a few casual remarks on the beautiful and challenging aspects of non-linear structure formation.
Before developing the full relativistic perturbation theory, we will derive results in Newtonian perturbation theory. These results will be a good description of the full GR calculation within the Hubble radius, for non-relativistic matter.

10.1 Structure formation from gravitational instability

To analyze the formation of large-scale structure quantitatively, we will proceed in two stages. In this section, we will discuss structure formation using Newtonian theory. This will give correct results for matter on sub-Hubble scales, providing valuable insight. However, it cannot adequately describe evolution on scales larger than the horizon, nor can it correctly describe the evolution of radiation perturbations. We will discuss structure formation from a full relativistic formalism in the next section where we will also include a radiation fluid.

10.2 Continuity and Euler equations in an expanding universe.

We begin by writing down the fluid equations for a non-relativistic fluid with mass density ρ, pressure P and velocity $\mathbf{u}$. The conservation of mass implies the *continuity equation*:

$$\partial_t \rho + \nabla_{\mathbf{r}} \cdot (\rho \mathbf{u}) = 0 \,. \tag{10.2}$$

Balancing forces on a fluid element as it moves (or conserving momentum) gives the *Euler equation*:

$$\partial_t \mathbf{u} + \mathbf{u}.\nabla_{\mathbf{r}} \mathbf{u} = -\nabla_{\mathbf{r}} p/\rho - \nabla_{\mathbf{r}} \Phi \,. \tag{10.3}$$

Finally, we can write down Poisson's equation to determine the gravitational potential Φ:

$$\nabla_r^2 \Phi = 4\pi G \rho \,, \tag{10.4}$$

where ρ is the total energy density of all constituents A comoving observer in the background cosmology has the physical coordinate

$$\mathbf{r}(t) = a(t)\mathbf{x} \,, \tag{10.5}$$

where $\mathbf{x}$ is the comoving coordinate. Note that $\nabla_r = \frac{1}{a}\nabla$. Derivatives transform in the following manner:

$$\left(\frac{\partial}{\partial t}\right)_{\mathbf{r}} = \left(\frac{\partial}{\partial t}\right)_{\mathbf{x}} + \left(\frac{\partial \mathbf{x}}{\partial t}\right)_{\mathbf{r}} \cdot \nabla = \left(\frac{\partial}{\partial t}\right)_{\mathbf{x}} - H(t)\mathbf{x} \cdot \nabla \,. \tag{10.6}$$

Here ∇ is a gradient with respect to $\mathbf{x}$ taken at fixed t. In what follows, time derivatives should be assumed to be evaluated at fixed $\mathbf{x}$.
We now perturb the density about its background value:

$$\rho(\mathbf{x}, t) = \bar{\rho}(t) + \delta\rho(\mathbf{x}, t) = \bar{\rho}(t)\left[1 + \delta(\mathbf{x}, t)\right] \,. \tag{10.7}$$

We similarly perturb

$$p = \bar{p} + \delta p \,, \qquad \mathbf{u} = Ha\mathbf{x} + \mathbf{v} \,, \qquad \Phi = \bar{\Phi} + \phi \,. \tag{10.8}$$

Note that the background values do not depend on $\mathbf{x}$ but only on time. Finally, we will replace $\nabla_{\mathbf{r}} \to \frac{1}{a}\nabla$ and $\partial_t \to (\partial_t - H\mathbf{x} \cdot \nabla)$ to discuss physics in an expanding universe with comoving coordinates.
We begin by analyzing the continuity equation when perturbed in this manner. Inserting the above expressions we obtain:

$$\left[\partial_t - H\mathbf{x} \cdot \nabla\right]\left(\bar{\rho}(1+\delta)\right) + \frac{1}{a}\nabla \cdot \left[\bar{\rho}(1+\delta)(Ha\mathbf{x} + \mathbf{v})\right] = 0. \tag{10.9}$$

We may now expand this order by order.
At zeroth order in the fluctuations δ and $\mathbf{v}$, we can make use of the fact that spatial derivatives of background quantities vanish as well as the fact that $\nabla \cdot \mathbf{x} = 3$. Hence we obtain:

$$\frac{\partial \bar{\rho}}{\partial t} + 3H\bar{\rho} = 0 \tag{10.10}$$

This can be recognized as the homogeneus continuity equation for matter.

Isolating terms at first order in the perturbations, we obtain

$$[\partial_t - H\mathbf{x}\cdot\nabla]\,(\bar\rho\delta) + \frac{1}{a}\nabla\cdot[\bar\rho\delta Ha\mathbf{x} + \bar\rho\mathbf{v}] = 0. \tag{10.11}$$

Using again using $\nabla\cdot\mathbf{x} = 3$, this can be rearranged to:

$$[\dot{\bar\rho} + 3H\bar\rho] + \bar\rho\dot\delta + \frac{1}{a}\bar\rho\nabla\cdot\mathbf{v} = 0. \tag{10.12}$$

Since the first term is zero by the continuity equation, we then obtain

$$\dot\delta = -\frac{1}{a}\nabla\cdot\mathbf{v}\,. \tag{10.13}$$

We now turn to the Euler equation. Including perturbations and writing the equation in comoving coordinates as before we obtain:

$$[\partial_t - H\mathbf{x}\cdot\nabla]\,(Ha\mathbf{x}+\mathbf{v}) + (Ha\mathbf{x}+\mathbf{v})\cdot\frac{\nabla}{a}(Ha\mathbf{x}+\mathbf{v}) = -\frac{\nabla\delta p}{a\bar\rho(1+\delta)} - \frac{\nabla\phi}{a} \tag{10.14}$$

To zeroth order this equation is trivial because the background is isotropic. Isolating the linear terms as previously, we obtain after some rearrangement

$$\dot{\mathbf{v}} + H\mathbf{v} = -\frac{1}{a\bar\rho}\nabla\delta p - \frac{1}{a}\nabla\phi\,. \tag{10.15}$$

Straightforwardly, the Poisson equation becomes

$$\nabla^2\phi = 4\pi Ga^2\bar\rho\delta\,. \tag{10.16}$$

The three equations (10.13), (10.15) and (10.16) for a closed system of equations. We have two first order equations for the variables δ and $\mathbf{v}$ and a simple algebraic constraint that can be used to remove ϕ altogether. As discussed previously, we can SVT decompose these equations. Introducing $v_i = \omega_i + \partial_i\theta$ with the vorticity ω_i satisfying $\partial_i\omega_i = 0$ and the velocity divergence $\partial_i v_i = \theta$. We can now re-write the Euler equation as an equation for the scalar modes δ and θ simply by taking its divergence,

$$\dot\theta + H\theta = -\frac{1}{a\bar\rho}\nabla^2\delta p - \frac{1}{a}\nabla^2\phi\,. \tag{10.17}$$

The Newtonian limit of general relativistic equations While we have derived (10.13), (10.17) and (10.16) from scratch in this section, they actually correspond to the Newtonian limit of the relativistic equations we derived in Sec. 8.5. For simplicity we will show this only for the case of a non-relativistic fluid with negligible pressure. First, for the continuity equation, take (8.70) and substitute $\delta\rho = \bar\rho\delta$, set $\Phi = \Psi$ and drop all time derivatives of the Newtonian potentials, since only spatial derivatives survive in the Newtonian limit. Then set $p \sim 0$ as appropriate for a non-relativistic fluid with $w \simeq 0$ and use (10.10). This results precisely in (10.13) with the identification $\nabla^2\delta u/a^2 = \nabla\cdot\mathbf{v}/a$. These factors of a can be understood as follows. Recall that $u_i = \partial_i\delta u$ and so $\nabla^2\delta u/a^2 = \partial_i g^{ij}u_j$ is the physical velocity divergence, computed with the metric. Then $\mathbf{v}$ is the physical velocity, but ∇ is a derivative with respect to the comoving coordinate $\mathbf{x}$ and so the physical velocity divergence must be $\nabla \cdot \mathbf{v}/a$. Second, for the Euler equation, we apply the same approximations to (8.71). We set $\bar{p} = \delta p = 0$, use (10.10) and find exactly (10.17) with the identification $\partial_i \delta u = u_i = a v_i$. Third, for the Poisson equation we start either from the 00 or the spatial-trace part of the Einstein equations in (8.66) or (8.69). From either equation we drop all time derivatives of the metric potentials. Since on sub-Hubble scales $\nabla^2 \sim -k^2 \gg H^2, \dot{H}$ we can also drop the terms without any derivatives $\Phi = \Psi$. The only two terms surviving are precisely those in the Poisson equation (10.16) where we used $4\pi G_N = 1/(2M_{\rm Pl}^2)$.

The Jeans instability Now use the equations we have derived to compute the evolution of matter perturbations and hence we will write δ_m instead of simply δ. By "matter" for the moment we simply mean a substance with negligible pressure and hence a vanishing speed of sound. We will see shortly that this applies always to dark matter and to baryonic matter only after hydrogen recombination.

We may now combine the three scalar equations. Using (10.16) to remove ϕ, and (10.13) to substitute $\theta = \nabla \mathbf{v} = -a\dot{\delta}_m$ inside (10.17) we find

$$\ddot{\delta}_m + 2H\dot{\delta}_m - \frac{1}{a^2\bar{\rho}} \nabla^2 \delta_m p - 4\pi G \bar{\rho} \delta_m = 0\,. \tag{10.18}$$

This is the fundamental equation describing the growth of structure in Newtonian theory. It shows the competition between pressure support, the $\nabla^2 \delta p$ term which opposes collapse, and gravitational attraction, the $4\pi G\bar{\rho}\delta_m$ term, which favors collapse.

We will now consider the case of a barotropic fluid for which $p = p(\rho)$. Then

$$\delta p = \frac{\partial p}{\partial \rho} \bar{\rho} \delta_m = c_s^2 \bar{\rho} \delta_m\,, \tag{10.19}$$

where c_s is the speed of sound. Replacing the relevant pressure term and going to Fourier space gives

$$\ddot{\delta}_m + 2H\dot{\delta}_m + \left[\frac{c_s^2}{a^2} k^2 - 4\pi G\bar{\rho}\right] \delta_m = 0\,. \tag{10.20}$$

We note that this implies different behavior depending on wavenumber k, since the term in brackets then changes sign.

The critical wavenumber involved is the Jeans' wavenumber:

$$k_J \equiv \frac{\sqrt{4\pi G \bar{\rho} a^2}}{c_s}\,. \tag{10.21}$$

On scales smaller than the Jeans' wavenumber, i.e. for large k, the system is effectively a damped oscillator with a Hubble friction term; due to the pressure support, the fluid hence supports acoustic oscillations. a.k.a. sound waves. On larger scales, they grow as a power law. Equivalently, perturbations with proper wavelength larger than the *Jeans' scale* $\lambda_J = 2\pi a/k_J$ for

$$\lambda_J = c_s \sqrt{\frac{\pi}{G\bar{\rho}}} \tag{10.22}$$

are gravitationally unstable; in contrast, shorter perturbations undergo oscillations supported by the restoring force of pressure.

This result is supported by the following intuition. By considering the motion of the surface of a sphere collapsing under its own gravity, it can be shown that the time for a perturbation to freely collapse, i.e. the "free-fall time", is $t_f \sim 1/\sqrt{G\rho}$; the time for a sound wave to cross a perturbation of size R is $t_{sc} = R/c_s$. These two quantities are equal when $R \sim c_s\sqrt{\frac{1}{G\bar{\rho}}} \sim \lambda_J$. For perturbations larger than the Jeans' length ($R > \lambda_J$), pressure cannot successfully resist gravitational infall, since there is not sufficient time for an pressure wave, opposing infall, to respond to the disturbance before the perturbation has collapsed.

10.3 Dark matter evolution inside the Hubble radius

We now calculate the growth of dark matter perturbations inside the Hubble radius using Newtonian perturbation theory.

Matter domination During dark matter domination, we can neglect all other perturbations aside from those of matter, and similarly assume the Hubble parameter H is only determined by matter density via $3H^2M_{\rm Pl}^2 = \bar{\rho}_m \propto a^{-3}$. The speed of sound for a dark matter fluid can additionally be assumed to be zero. This is because dark matter appears to behave as collision-less particles and baryon density is too low after redshift $z \sim 1100$ for collisions to be important on large scales. It is only when baryons start to form very overdense regions that collisions and hence a speed of sound need to take into account.
We thus obtain from Equation (10.20):

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G\bar{\rho}_m\delta_m = 0 \tag{10.23}$$

Since during matter domination $a = (t/t_0)^{2/3}$, $H = 2/(3t)$ and hence (using also $H^2 = \frac{8\pi G\bar{\rho}}{3}$)

$$\ddot{\delta}_m + \frac{4}{3t}\dot{\delta}_m - \frac{2}{3t^2}\delta_m = 0. \tag{10.24}$$

Inserting a power law ansatz $\delta_m \propto t^c$, we obtain two solutions. The first is $\delta_m \propto 1/t$, which decays. The second is

$$\delta_m \propto \left(\frac{t}{t_0}\right)^{2/3} = a\,, \tag{10.25}$$

which grows linearly in a. We will therefore ignore the second solution in the following. The growing mode solution describes the growth of matter perturbations linearly in a during matter domination inside the Hubble radius. This is a key result in structure formation and a good approximate description of our universe during $3300 \gg z \gg 1$. This solution is relevant to describe the growth of dark matter inhomogeneities at all times during matter domination, i.e. after $z = 3300$. This solution applies to baryon as well but only after they decouple from photons and become collision-less, which happens after recombination at $z = 1100$.

Radiation domination Now we look at the same equation but we assume that the total energy density of the universe is dominated by radiation, comprising of photons and relativistic neutrinos, $\rho_r = \rho_\gamma + \rho_\nu$. Therefore,

$$3H^2M_{\rm Pl}^2 = \bar{\rho}_r \propto a^{-4}\,, \tag{10.26}$$

in which case the expansion is faster than in matter domination. This approximation is valid in our Universe before $z_{\rm eq} = 3300$, corresponding to matter radiation equality, $\rho_m(z_{\rm eq}) = \rho_r(z_{\rm eq})$. During radiation domination, since it is the total energy density that sources the potential, we have to use the Poisson equation

$$\nabla^2\phi = 4\pi G_N a^2(\bar{\rho}_{dm}\delta_{dm} + \bar{\rho}_b\delta_b + \bar{\rho}_\gamma\delta_\gamma)\,. \tag{10.27}$$

Here we neglected neutrino perturbations neutrino perturbations because they decay quickly upon entering the Hubble radius. On sub-Hubble scales, $k \ll aH$, we will show later that the perturbations δ_γ of the photon density ρ_γ oscillate rapidly, with sound waves sustained by a significant radiation pressure. Since on timescales of structure growth this oscillation averages to zero, to first approximation we can simply neglect the radiation density perturbation (this will be shown in more detail later using relativistic perturbation theory). Neutrino perturbations are also effectively zero as they decay quickly upon entering the Hubble radius. Finally, since radiation domination happened before recombination, the electrons and protons were free charged particles and interacted frequently with photons forming effectively a single charged fluid, the electron-baryon-photon plasma. Because of the relativistic nature of photons this fluid had a pressure comparable to the energy density and so a correspondingly large speed of sound. As a consequence baryon don't collapse at all during this time and their overdensities do not grow. Let's then focus on just dark matter inhomogeneities δ_{dm}.
In radiation domination $a = (t/t_0)^{1/2}$ and hence $H = \frac{1}{2t}$. The equation for the evolution of matter density perturbations then becomes

$$\ddot{\delta}_{dm} + \frac{1}{t}\dot{\delta}_{dm} - 4\pi G\bar{\rho}_{dm}\delta_{dm} = 0\,, \tag{10.28}$$

The exact solutions of this equation are Bessel functions. For an approximate solution we note that, since the evolution of the density perturbation must depend on the Hubble rate (there is no other timescale in the absence of pressure), we should expect $\ddot{\delta}_{dm} \sim H^2\delta_{dm} \sim 8\pi G\bar{\rho}_r\delta_{dm}/3 \gg 4\pi G\bar{\rho}_{dm}\delta_{dm}$, because by assumption we are in radiation domination, $\bar{\rho}_r \gg \bar{\rho}_m$. A similar estimate suggests that $\frac{1}{t}\dot{\delta}_{dm} \gg 4\pi G\bar{\rho}_{dm}\delta_{dm}$. We may therefore drop the third term in this equation, obtaining

$$\ddot{\delta}_{dm} + \frac{1}{t}\dot{\delta}_{dm} \simeq 0\,.$$

One obvious solution is $\dot{\delta}_{dm} = 0$ and hence δ_{dm} constant. The second is obtained by first integrating once to find $\dot{\delta}_{dm} = 1/t$, and hence integrating again to find $\delta_{dm} \propto \ln t$. This second solution is growing and hence leading over the first

$$\delta_{dm} \propto \ln a, \tag{10.29}$$

We conclude that early on, in radiation domination, dark matter density fluctuations grow extremely slowly within the horizon.

Dark energy domination In our Universe the dark energy dominated era starts after matter-dark energy equality at redshift $z_{m\Lambda}$ satisfying

$$\Omega_{m0}(1+z_{m\Lambda})^3 = \Omega_{\Lambda 0} \quad \Rightarrow \quad z_{m\Lambda} = \left(\frac{0.7}{0.3}\right)^{1/3} - 1 \simeq 0.3\,. \tag{10.30}$$

By this time we can completely neglect radiation and its perturbations since

$$\Omega_r = \Omega_{r,eq} \frac{1+z_{m\Lambda}}{1+z_{eq}} \sim 10^{-4} \, .$$

Moreover, a cosmological constant has no inhomogeneities as it is constant both in time and in space. We conclude that the only perturbations sourcing ϕ in the Poisson equations are those of matter, both baryonic and dark. Hence we find again

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G\bar{\rho}_m \delta_m = 0 \, . \tag{10.31}$$

However, Λ is still highly relevant for this equation, since it strongly affects the Hubble rate. Let's start deriving an approximated solution deep into a dark energy dominated cosmology, which will happen in the future for us. For the Hubble parameter we have

$$H^2 \approx \frac{8\pi G\bar{\rho}_\Lambda}{3} \approx \text{const} \gg 4\pi G\bar{\rho}_m \, .$$

As previously for radiation, in this limiting case we can drop the final term to obtain:

$$\ddot{\delta}_m + 2H\dot{\delta}_m \simeq 0 \tag{10.32}$$

which, with H constant, gives solutions:

$$\delta_m \propto \text{const} \, . \tag{10.33}$$

as well as a decaying solution $\propto e^{-2Ht} \propto a^{-2}$. We conclude that deep in the dark energy dominated era, perturbations stop growing. For a more accurate solution we have to solve the equations numerically and the result is shown in Fig. **??**. We see the linear in a growth during matter domination transitioning to the to constant δ_m solution in the future. Since dark energy domination has started only recently, δ_m is still slightly growing today, but slower than linear in a.

This result is a general feature of any constituent that contributes to the total energy density, and hence to H, but does not contribute to ϕ by clustering: the growth of structures is suppressed compared to matter domination. Massive neutrinos, for example, have minimal fluctuations below their free-streaming scale, and hence contribute only to H but not to δ_m on small scales, thus reducing small-scale structure growth. Cosmological experiments aiming to precisely determine the growth of cosmic structure are hence expected to be very powerful: from the details of how structure grows, we can place strong constraints on the properties of the different components of our universe.

11 Large scale structures

In this section, we will evaluate the evolution of cosmological perturbations beyond the range of validity of Newtonian perturbation theory: using relativistic perturbation theory, we will calculate the evolution not just of matter, but also of Newtonian potentials and radiation, thus deriving the growth of structure over a wide range of times and scales.
We will make use of the formalism for cosmological perturbation theory that we have developed and the evolution equations that we have derived in the previous section.
We begin by reviewing an important concept for clarity: that of *sub-Hubble* and *super-Hubble* scales. As the comoving Hubble radius $1/(aH)$ grows with time after the end of inflation, since now $aH = \dot{a}$ decreases during decelerated expansion until very recently, modes with larger and larger comoving wavelength k re-enter our Hubble patch. At a given time, a mode with a conformal wavelength k^{-1} that is larger than the conformal Hubble radius $1/(aH)$, i.e.

$$\text{super-Hubble mode:} \qquad k \ll (aH)\,.$$

After horizon entry, perturbations become

$$\text{sub-Hubble mode:} \qquad k \gg 1/(aH)\,.$$

From our study of adiabatic modes we know that the adiabatic mode, which is the only relevant scalar perturbation in our universe, is constants on super-Hubble scales, when appropriately defined in terms of gauge invariant variables such as $\mathcal{R}$ or ζ. It should be mentioned though that this is a peculiarity of this solution. Other, non-adiabatic modes would evolve also on super-Hubble scales.
To continue our discussion, we first remind the reader of the key evolution equations we have derived in the previous section: the relevant equations are the continuity equation (8.71), the Euler equation (8.70) and the Einstein equations (8.66)-(8.69).
It will be convenient to proceed using conformal time τ as opposed to cosmological time t. We denote ∂_t with a dot and ∂_τ with a prime. The Friedmann equations in conformal time becomes

$$\mathcal{H}^2 = \frac{8\pi G\bar{\rho}a^2}{3}\,, \tag{11.1}$$

where we have defined the convenient *conformal Hubble parameter* $\mathcal{H} = aH$. Some useful formulae for conversion are

$$\mathcal{H} = \frac{a'}{a} = aH = \dot{a} \qquad H = \frac{\dot{a}}{a} = \frac{\mathcal{H}}{a}\,, \qquad \frac{\ddot{a}}{a} = \dot{H} + H^2 = \frac{\mathcal{H}'}{a^2}\,, \tag{11.2}$$

$$\dot{H} = \frac{1}{a^2}\left(\mathcal{H}' - \mathcal{H}^2\right)\,, \qquad \partial_t = \frac{1}{a}\partial_\tau\,, \qquad \partial_t^2 = \frac{1}{a^2}\left(\partial_\tau^2 - \mathcal{H}\partial_\tau\right)\,. \tag{11.3}$$

In the presence of a single fluid with equation of state w the Friedman and continuity equations give the relation

$$2\mathcal{H}' + \mathcal{H}^2(3w+1) = 0\,. \tag{11.4}$$

In the special case in which only one barotropic component dominates the energy density of the Universe one can combine the Einstein equations in Newtonian gauge to derive a relatively simple equation for Φ alone. To this end, note that there is one equation of state $p = w\rho$ and so we can write $\delta p = w\delta\rho$ in (8.66). Then we solve for $\delta\rho$ and plug it into (8.69), to obtain:

$$\Phi'' + 3(1+w)\mathcal{H}\Phi' + \left[2\mathcal{H}' + \mathcal{H}^2(3w+1)\right]\Phi + k^2 w\Phi = 0. \tag{11.5}$$

where we moved to conformal time. We can simply this further using the background solution in (11.4),

$$\Phi'' + 3(1+w)\mathcal{H}\Phi' + k^2 w\Phi = 0 \tag{11.6}$$

Unfortunately this equation is not valid during the transition from radiation to matter domination or from matter to dark energy because more than one fluid is relevant.

11.1 Super-Hubble limit and initial conditions

As we know from our discussion of adiabatic modes, the initial condition for all perturbations are most simply described in terms of the gauge invariant comoving curvature perturbation $\mathcal{R}$. At the end of inflation, all relevant modes $\mathbf{k}$ are assumed to be outside the Hubble radius, $k \ll \mathcal{H}$ and in the adiabatic mode. Since we will solve our equations in Newtonian gauge, we recall that the Newtonian potentials $\Psi = \Phi$ are related to $\mathcal{R}$ by (8.101). In the single-fluid approximation we derived the relation (8.106), namely

$$\mathcal{R} = -\frac{5+3w}{3+3w}\Phi \tag{11.7}$$

In this approximation is it easy to check that indeed $\mathcal{R}$ and Φ are constant on super-Hubble scales. Since $k \ll \mathcal{H}$, we can neglect the $k^2 w\Phi$ term in (11.6), to find

$$\Phi'' + 3(1+w)\mathcal{H}\Phi' = 0. \tag{11.8}$$

This has solution

$$\Phi = \text{const.} \quad (k \ll aH) \tag{11.9}$$

as well as a second decaying solution we neglect. This is true as long as w is a constant (i.e., the universe is dominated by one component), and is hence true in both the radiation and matter eras. However, because w changes when radiation-domination turns to matter-domination, the potentials differ in these two eras. To compute the change we will connect the potentials to $\mathcal{R}$ before and after matter radiation equality and invoke the super-Hubble conservation of $\mathcal{R}$ which we proved for general consmologies.

Using $w = 1/3$ and $w = 0$ for radiation and matter domination we can relate the conserved curvature perturbation to the Newtonian potential in radiation domination Φ_{RD} and in matter domination Φ_{MD}:

$$\mathcal{R} = -\frac{3}{2}\Phi_{RD} = -\frac{5}{3}\Phi_{MD}\,. \tag{11.10}$$

We can make use of the conservation of $\mathcal{R}$ to conclude that $\Phi_{MD} = \frac{9}{10}\Phi_{RD}$, i.e. the potentials decrease slightly as the Universe goes from radiation domination to matter domination. We caution the reader that it is generally wise not to attribute much physical meaning to the behavior gauge dependent perturbations such as Φ on super-Hubble scales.

From our determination of the potential evolution on super-horizon scales, we can derive the behavior of super-Hubble radiation and matter perturbations. We already know that all scalar perturbations on super-Hubble scales are given by (8.102), which reduces to (8.106) for single-fluid cosmologies. For density perturbations to constituent a, $s = \rho_a$, we can massage (8.106) into

$$H\frac{\delta\rho_a}{\dot{\bar{\rho}}_a} = \frac{\delta\rho_a}{-3(\bar{\rho}_a + \bar{p}_a)} = -\frac{\delta_a}{3(1+w_a)} = \Phi\frac{2}{3(1+w)}\,, \tag{11.11}$$

Moreover, if constituent a is also the dominant energy density, so that $w = w_a$, then $\delta = -2\Phi$. This simple expression holds for perturbations to the dominant substance in the Universe. The others substances can be related to it using

$$\frac{\delta_a}{(1+w_a)} = \frac{\delta_b}{(1+w_b)} \qquad \text{for all } a \text{ and } b\,.$$

Let's check that this is indeed the right solution of the equations of motion for a single fluid. Going back to (8.66) and (8.69) we can drop all $\dot\Phi$ and $\ddot\Phi$ terms because Φ is constant and all $\nabla^2\Phi$ terms because we are on super-Hubble scales. We then combine the two equations to solve separately for $\delta\rho$ and δp. As expected we obtain

$$\delta = -2\Phi = \text{const.} \quad (\text{super-Hubble }, k \ll \mathcal{H}, \delta \text{ dominant})\,. \tag{11.12}$$

In radiation domination, assuming adiabatic perturbations, we hence obtain for super horizon scales:

$$\delta_r = \frac{4}{3}\delta_m = -2\Phi_{RD} = \text{const.} \quad (\mathrm{super-Hubble, k \ll \mathcal{H}}) \tag{11.13}$$

In matter domination, we obtain that:

$$\delta_m = \frac{3}{4}\delta_r = -2\Phi_{MD} \tag{11.14}$$

since the matter perturbation then sources the potential.
Having described the (constant) behavior on super horizon scales, we will now examine the evolution of (i) fluctuations in the potential, (ii) in the radiation, and (iii) in the matter once they enter the Hubble radius and evolve. In each case we will consider the evolution both in radiation domination and in matter domination. For (iii) we will also discuss the current era, when dark energy has become important.

11.2 Potential Evolution

We first examine the evolution of the gravitational potential. Our basic equation for this case will be (11.6). For both radiation and matter domination we must simply solve this equation for $w = 1/3$ and $w = 0$ respectively; we must also recall that

$$\text{radiation domination: } \mathcal{H} = \frac{1}{\tau}\,, \qquad a = \frac{\tau}{\tau_0}\,, \tag{11.15}$$

$$\text{matter domination: } \mathcal{H} = \frac{2}{\tau}\,, \qquad a = \left(\frac{\tau}{\tau_0}\right)^2\,. \tag{11.16}$$

Radiation Domination We begin by discussing radiation domination ($w = 1/3$). Equation (11.6) then becomes:

$$\Phi'' + \frac{4}{\tau}\Phi' + \frac{k^2}{3}\Phi = 0\,. \tag{11.17}$$

The solution to this equation is given in terms of spherical Bessel and Neumann functions (j_1, n_1):

$$\Phi(\mathbf{k}) = A(\mathbf{k})\frac{j_1(k\tau/\sqrt{3})}{k\tau/\sqrt{3}} + B(\mathbf{k})\frac{n_1(k\tau/\sqrt{3})}{k\tau/\sqrt{3}}\,, \tag{11.18}$$

which are defined as

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x}; \qquad n_1(x) = -\frac{\cos x}{x^2} - \frac{\sin x}{x}\,. \tag{11.19}$$

These functions have the low-x expansion

$$j_1(x) \approx \frac{x}{3} + O(x^2)\,, \qquad\qquad n_1(x) \approx -\frac{1}{x^2} + O(x^0)\,. \tag{11.20}$$

This implies that for early times ($x \to 0$), the $B_{\mathbf{k}}$ term (describing a decaying mode) becomes singular – matching to inflationary perturbations implies we must set it to zero. We are left with only the $A(\mathbf{k})$ term. In the early time (low x) limit, we therefore obtain $\Phi \approx A/3$. Setting $\mathcal{R} = -\frac{3}{2}\Phi$ and solving for A, we obtain $A = -2\mathcal{R}$ and hence

$$\Phi(\tau, \mathbf{k}) = -2\mathcal{R}(\mathbf{k})\frac{\sin(k\tau/\sqrt{3}) - k\tau\cos(k\tau/\sqrt{3})/\sqrt{3}}{(k\tau/\sqrt{3})^3}\,. \tag{11.21}$$

We can now deduce the behavior during radiation domination, noting that $\tau = 1/\mathcal{H}$ is the comoving horizon size. For early times when $\tau = 1/\mathcal{H} \ll 1/k$, when the mode is still outside the horizon, the mode is constant (and $= -2/3\mathcal{R}$). Well inside the horizon, when $k\tau \gg 1$, we obtain the following evolution:

$$\Phi(\tau, \mathbf{k}) = 6\mathcal{R}(\mathbf{k})\frac{\cos(k\tau/\sqrt{3})}{(k\tau)^2}\,. \tag{11.22}$$

We conclude: when the Φ mode enters the horizon, i.e. when $\tau = 1/\mathcal{H} > 1/k$, the mode begins to oscillate; the oscillations are multiplied by a decaying envelope, with their amplitude falling as $1/\tau^2$.

Matter Domination During matter domination, we can take $w = 0$. The equation describing the potential evolution (11.6) now becomes (recalling that $\mathcal{H} = 2/\tau$):

$$\Phi'' + \frac{6}{\tau}\Phi' = 0\,. \tag{11.23}$$

The growing mode solution to this is a simple constant

$$\Phi = \text{const.} \quad \text{(all scales)}\,. \tag{11.24}$$

The evolution of a mode as a function of cosmic time thus depends on its wavelength compared to the horizon size at matter-radiation equality, or equivalently k compared to $k_{eq} \simeq 10^{-2}\text{Mpc}^{-1}$:

- If $k \ll k_{eq} = \mathcal{H}$, i.e. we have a very large mode that only enters the horizon at late times after matter radiation equality, it has been nearly constant during the radiation era. Its amplitude is nearly preserved during the matter era, with only a suppression of 9/10 from the radiation-to-matter transition.
- If $k \gg k_{eq}$, the mode oscillates during the radiation era; its amplitude is also reduced by the oscillation envelope by $\propto a^{-2}$ or (substantially if $k >> k_{eq}$). The small amplitude is again preserved during the matter era.
- If $k \approx k_{eq}$, a small amount of decay happens as the horizon scale approaches the mode size during the radiation era, though this decay halts during the matter era.

The potential evolution in these three scenarios is illustrated in Fig. 22.

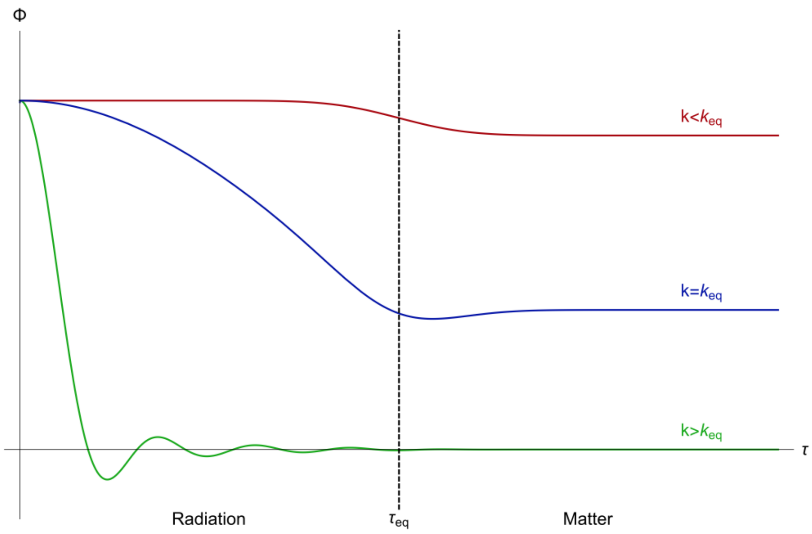


Figure 22: Evolution of the potential as a function of time for a mode that enters the horizon after matter-radiation equality (red line), at matter-radiation equality (blue) and well before matter-radiation equality, during the radiation era (green). It can be seen that inside the horizon, during radiation domination, the potentials oscillate and decay; otherwise, well outside the horizon or within matter domination, the potentials stay constant. (Image credit: J. Fergusson.)

11.3 Radiation Evolution

We now discuss the evolution of radiation perturbations. Using $\Phi = \Psi$, we can write δu as function of Φ from (8.72). We substitute the result into the relativistic Poisson equation (8.74) to find

$$\delta = -\frac{2}{3}\frac{k^2\Phi}{\mathcal{H}^2} - 2\Phi'/\mathcal{H} - 2\Phi',, \tag{11.25}$$

where δ are the perturbations to the dominant source of energy density, namely δ_r during radiation domination and δ_m during matter domination. Given a solution for Φ, we can use this equation to deduce the evolution of the perturbations to the leading source of energy density. Intuitively this relation arises because the Newtonian potentials are determined by the inhomogeneities in the distribution of all forms of matter. Inverting this relation, the knowledge of Φ tells us about the inhomogeneity. However, in the presence of multiple fluid, gravity can only tell us the total inhomogeneities, which are dominated by the dominant source of energy density at the time. To find out perturbations to subleading constituents, such as matter during radiation domination, we will need to separately solve the corresponding continuity and Euler equations. Let's see what the above relation gives for radiation and matter domination respectively.

Before getting into the details, we note that we have a choice of two variables with which to evaluate a density perturbation: we can use either δ or $\Delta = \delta - 3\mathcal{H}(1+w)v$. However, both variables are supposed to describe a physical quantity, the density contrast, in different gauges – which is the one we actually can measure? The answer is that they are both valid to use and both will give correct results, because they are both identical on sub-horizon scales, where we can actually make observations. To see this, note that $\mathbf{v} = \nabla v$, so that in Fourier space

$$\Delta - \delta = 3\mathcal{H}(1+w)v \approx 3\mathcal{H}(1+w)|\mathbf{v}|/k \propto \frac{\mathcal{H}}{k}|\mathbf{v}| \approx 0 \tag{11.26}$$

where $\frac{\mathcal{H}}{k} \approx 0$ on sub-Hubble scales. We may thus use whichever variable is most convenient to perform calculations of observables.

Radiation Domination In radiation domination, the Poisson equation becomes

$$\delta_r = -\frac{2}{3}k^2\tau^2\Phi - 2\Phi'\tau - 2\Phi\,. \tag{11.27}$$

Well outside the horizon ($\tau \ll 1/k$), we deduce that δ_r is constant since Φ is constant, which we knew from our general discussion around (11.13). Inside the Hubble radius, $\tau k \gg 1$ the last two terms are subleading compared to $k^2\tau^2\Phi$ and we drop them to find

$$\delta_r \approx -\frac{2}{3}(k\tau)^2\Phi = -4\mathcal{R}(0)\cos\left(\frac{1}{\sqrt{3}}k\tau\right). \tag{11.28}$$

We conclude that on sub horizon scales, the fluctuations in the radiation density δ_r undergo sinusoidal oscillations around a point $\delta_r = 0$.

Matter Domination During matter domination, radiation is a subleading form of energy density and so its inhomogeneities cannot be read off directly from the Newtonian potentials. Instead, we must return to the continuity and Euler equations for radiation, (8.71) and (8.70). Using $\bar{p}_r = \bar{\rho}_r/3$ and $\delta p_r = \bar{\rho}_r\delta_r/3$, these become

$$\dot{\delta u_r} - H\delta u_r + \frac{1}{4}\delta_r + \Phi = 0\,, \qquad \dot{\delta}_r - \frac{4}{3}\frac{k^2}{a^2}\delta u - 4\dot{\Phi} = 0\,, \tag{11.29}$$

We can now use that the Newtonian potential Φ is constant in matter domination to drop $\dot{\Phi}$. Then we go to comoving time and solve the second equation for δu_r

$$\delta u_r = \frac{a}{k^2}\frac{3}{4}\delta'\,. \tag{11.30}$$

Substituting this into the first equation we find a second order equation in time for δ_r alone

$$\delta_r'' - \frac{1}{3}\nabla^2\delta_r = \frac{4}{3}\nabla^2\Phi = \text{const.} \tag{11.31}$$

Recalling that the potential is constant during matter domination, we can also rewrite it as

$$(\delta_r(k,\tau)/4 + \Phi(k))'' + \frac{1}{3}k^2(\delta_r(k,\tau)/4 + \Phi(k)) = 0\,. \tag{11.32}$$

This is driven harmonic oscillator for each k-mode – showing the same oscillations as in radiation domination, but now oscillating about a new equilibrium point $\delta_r = -4\Phi_{MD}(k)$, where the constant $\Phi(k)$ in matter domination is described in the previous discussion.
These oscillations in the radiation density are the so-called acoustic oscillations, which we can observe in the CMB and in the Baryon Acoustic Oscillation (BAO) feature in large-scale structure. The physics of these oscillations is simple: a radiation perturbation collapses due to gravity, but then radiation pressure exerts a restoring force, driving the perturbation back to a more dilute state; the process then repeats, giving oscillations.

11.4 Matter Evolution

We now consider the evolution of sub-horizon fluctuations in dark matter across cosmic history. We will consider radiation domination, matter domination, and finally the more recent era when both dark energy and matter are important. Since dark matter is about seven times more abundant than baryons, we will ignore baryons in our analysis as a first approximation.

Sub-Hubble evolution: Radiation and Matter Domination Our starting point are the continuity and Euler equations, (8.71) and (8.70), for matter in Newtonian gauge. Specifying to matter by taking $p \approx 0$ the continuity and Euler equations become

$$\dot{\delta}_m + \frac{\nabla^2}{a^2}\delta u - 3\dot{\Phi} = 0\,, \qquad \dot{\delta u} + \Phi = 0\,. \tag{11.33}$$

In Fourier space we can solve the first equation for δu to find

$$\delta u_m = -\frac{a^2}{k^2}\left(3\dot{\Phi} - \dot{\delta}_m\right)\,. \tag{11.34}$$

Plug this into the second equation and moving to conformal time this becomes

$$\delta_m'' + \mathcal{H}\delta_m' = \nabla^2\Phi + 3(\Phi'' + \mathcal{H}\Phi')\,. \tag{11.35}$$

On sub-Hubble scales only the first term on the right-hand side contributes, while the time derivatives of Φ are subleading. We have discussed earlier that the potential fluctuations oscillate rapidly on sub-Hubble scales, during radiation domination. We can think of the potential being approximately divisible into a radiation and matter component $\Phi = \Phi_r + \Phi_m$. As we saw previously, the radiation component is oscillatory, whereas the matter part is approximately constant. While the radiation can rapidly respond to potential fluctuations, the oscillations are too rapid to couple strongly to the matter perturbations. We can thus neglect the radiation fluctuations; the time averaged gravitational potential is then only sourced by the matter fluctuations $\Phi \approx \Phi_m$, with $\Phi'_m \approx \Phi''_m \approx 0$. The relativistic Poisson equation in (8.74) now reads

$$\delta_m - 3H\delta u_m = \frac{1}{4\pi G_N}\frac{\nabla^2}{a^2}\Phi\,. \tag{11.36}$$

From (11.34) we know that on sub-Hubble scales,

$$H\delta u \sim (a/k)^2 H\dot{\delta}_m \sim (aH/k)^2\delta_m \ll \delta_m\,,$$

and so we can drop the second term on the left-hand side of (11.36). Using this result to remove $\nabla^2\Phi$ in (11.35) we find

$$\delta''_m + \mathcal{H}\delta'_m - 4\pi G\bar{\rho}_m a^2\delta_m = 0. \tag{11.37}$$

From the Friedmann equation we can write the relevant conformal Hubble parameter as

$$\mathcal{H}^2 = \frac{H_0^2\Omega_{m0}^2}{\Omega_{r0}}\left(\frac{1}{y} + \frac{1}{y^2}\right)\,, \tag{11.38}$$

where $y = a/a_{eq}$ and we used $a_{eq} = \Omega_{r0}/\Omega_{m0}$.
We can now re-write (11.37) as the so-called Meszaros equation:

$$\frac{\mathrm{d}^2\delta_m}{\mathrm{d}y^2} + \frac{2+3y}{2y(1+y)}\frac{\mathrm{d}\delta_m}{\mathrm{d}y} - \frac{3}{2y(1+y)}\delta_m = 0\,. \tag{11.39}$$

The two solutions are

$$\delta_m \propto 2+3y\,, \qquad \delta_m \propto (2+3y)\ln\left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right) - 6\sqrt{1+y}\,. \tag{11.40}$$

During radiation, i.e. for $y \ll 1$, this implies that the growing mode is $\propto \ln y \propto \ln a$. During matter domination, i.e. for $y \gg 1$, the growing mode is $\delta_m \propto a$ as expected.
The matter dominated result can also be derived directly from the gravitational potential solution (which is constant after matter radiation equality), using the Poisson equation for the growing mode

$$\delta_m = \frac{\nabla^2\Phi}{4\pi G a^2\bar{\rho}} \propto a \tag{11.41}$$

We conclude that during radiation domination, the sub-Hubble growth nearly halts, $\delta_m \propto \log(a)$, while during matter domination, perturbations grow as $\delta_m \propto a$.

Dark Energy Becomes Important First, we note that at late times we have

$$\nabla^2\Phi = 4\pi G a^2\bar{\rho}_m\delta_m \tag{11.42}$$

since there are no fluctuations in the dark energy density. With $\bar{\rho} \propto a^{-3}$, this implies that for each wavenumber $\Phi \propto \delta_m/a$.
We now consider the Einstein equation for potential evolution. We can start from (11.5) and set $w = 0$ to represent the fact that matter has negligible pressure,

$$\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = 0. \tag{11.43}$$

We wish to re-arrange this into an equation for δ_m; we can do so by noting $\Phi \propto \delta_m/a$ and hence obtain:

$$(\delta_m/a)'' + 3\mathcal{H}(\delta_m/a)' + (2\mathcal{H}' + \mathcal{H}^2)(\delta_m/a) = 0. \tag{11.44}$$

This can be rewritten as

$$\delta_m'' + \mathcal{H}\Delta_m' + (\mathcal{H}' - \mathcal{H}^2)\delta_m = 0. \tag{11.45}$$

From the Friedmann equations, $\mathcal{H}' - \mathcal{H}^2 = -4\pi G a^2 \bar{\rho}_m$, and hence

$$\delta_m'' + \mathcal{H}\Delta_m' - 4\pi G\bar{\rho}_m a^2 \delta_m = 0. \tag{11.46}$$

This is the same equation as in the Newtonian case, and we obtain the same suppression of structure growth due to dark energy – i.e. the growth halts and the perturbation stays constant when the universe is entirely dark energy dominated.

11.5 Matter Power Spectrum

We are finally in the position to discuss the matter power spectrum. We start with a scale-invariant initial power spectrum of curvature perturbations, $P_{\mathcal{R}} \propto k^{-3}$. Now we separate the discussion for modes k that enter the Hubble radius either during matter domination or during radiation domination. The dividing wavenumber $k_{eq} \sim 10^{-2}\text{Mpc}^{-1}$ re-entered precisely at matter radiation equality.

$k < k_{eq}$**: large scales.** This case is particularly simple because we found that the Newtonian potential Φ is constant during matter domination both outside and inside the Hubble radius and moreover the appropriate initial condition is

$$\Phi_{MD} = -(3/5)\mathcal{R}\,. \tag{11.47}$$

Therefore the power spectrum of Φ is the same as that of $\mathcal{R}$, i.e. scale invariant. We can now wait until a mode is sufficient inside the Hubble radius that $k \gg aH$. These are actually the only modes we manage to observe in practice because on very large scales galaxy surveys are dominate by shot noise. On these sub-Hubble scales we can use the Poisson equation $-k^2\Phi(k) = 4\pi G a^2 \bar{\rho}\delta_m(k)$ to write

$$\delta_m(\mathbf{k}, t) = -\frac{k^2}{4\pi G_N a^2 \bar{\rho}_m}\Phi_{MD} \tag{11.48}$$

Putting everything together, the matter power spectrum, which is defined by

$$\langle \delta_m(\mathbf{k}, t)\delta_m(\mathbf{k}', t)\rangle = (2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}')P_{\delta_m}(k, t)\,, \tag{11.49}$$

is linearly proportional to the primordial power spectrum of curvature perturbations $\mathcal{R}$,

$$P_\delta = \left(\frac{2k^2}{5H_0^2\Omega_0}\right)^2 a^2 P_{\mathcal{R}}(k) \qquad (k < k_{eq})\,, \tag{11.50}$$

where we used

$$\bar{\rho}_m = \frac{\bar{\rho}_{m0}}{a^3} = 3H_0 M_{\text{Pl}}^2 \frac{\Omega_{m0}}{a^3}\,. \tag{11.51}$$

A few comments are in order. First, the relation (11.50) is valid to linear order in perturbation theory but for any primordial curvature power spectrum. For the specific choice of a scale invariant power spectrum $P_{\mathcal{R}} \propto k^{-3}$ corresponds to a large scale power spectrum that is linear in k. Indeed, the spectral tilt n_s was precisely introduced to parameterized deviations from this linear behaviour on large scales, $P_\delta \propto k^{n_s}$. Second, the time dependence comes exclusively from the scale factor, $P_\delta \propto a^2$. In particular the time and scale dependence factorize, i.e. $P_\delta(k_1,t)/P_\delta(k_2,t)$ does not depend on time.

$k > k_{eq}$**: short scales.** Now we consider modes that re-enter the Hubble radius during radiation domination. Our strategy will be to follow the evolution of the Newtonian potential Φ from Hubble re-entry throughout radiation domination and until matter domination when Φ become constant. Then, in matter domination we simply use (11.48) to relate Φ to δ_m, both of which are very sub-Hubble by then.
Around (11.22) we found that during radiation domination Φ decays as $\tau^{-2} \propto a^{-2}$. Therefore, by the time we reach matter radiation equality, each mode has decayed an amount $(a_k/a_{eq})^2$, where a_k is the value of the scale factor when the mode k re-entered the Hubble radius and started decaying, $a_k H_k = k$. Recalling that in radiation domination $H \propto \sqrt{\rho_r} \propto a^{-2}$ we can write

$$\left(\frac{a(k)}{a_{eq}}\right)^2 = \left(\frac{a_{eq}H_{eq}}{a_k H_k}\right)^2 = \left(\frac{k_{eq}}{k}\right)^2 . \tag{11.52}$$

One last effect that we missed so far is that matter perturbations grow logarithmically during radiation domination, as we found in (11.40). This contribution to the Newtonian potential was neglected in the calculation leading to (11.22), which only accounted for perturbation in the dominant radiation density. The end result is that Φ_{MD} is a larger than what is give above by a factor of $\log(a_{eq}/a(k))$, where again $a(k)$ is the scale factor of Hubble re-entry, when the logarithmic enhancement takes effect.
Including this logarithmic correction, we conclude that for $k > k_{eq}$, the matter power spectrum is given by

$$P_\delta = \left(\frac{2k^2}{5H_0^2\Omega_0}\right)^2 a^2 \frac{k_{eq}^2}{k^2} \log\left(\frac{k}{k_{eq}}\right)^2 P_{\mathcal{R}}(k) \qquad (k > k_{eq}) . \tag{11.53}$$

Sometimes a more precise expression for the matter power spectrum is required. For example one may want to take into account the smooth transition between radiation and matter domination, or the fact that the growth of structure slows down as dark energy comes into effect. To this end, one defines

$$P_\delta = \left(\frac{2k^2}{5H_0^2\Omega_0}\right)^2 D(t)^2 T(k/k_{eq}) P_{\mathcal{R}}(k) , \tag{11.54}$$

where $T(k/k_{eq})$ is known as the matter transfer function and $D(t)$ as the matter growth function. From our previous calculation we know the following approximate scalings:

$$k \ll k_{eq} : T(k/k_{eq}) \simeq 1 , \qquad k \gg k_{eq} : T(k/k_{eq}) \simeq \left[\frac{k_{eq}}{k} \log\left(\frac{k}{k_{eq}}\right)\right]^2 . \tag{11.55}$$

In summary, the matter power spectrum is given on larges scales by

$$P_\delta(k) \propto a^2\, k \qquad (k < k_{eq}) \tag{11.56}$$

and on short scales by

$$P_\delta(k) \propto a^2\, k^{-3}\, [\log(k/k_{eq})]^2 \qquad (k > k_{eq}) \tag{11.57}$$

This matches extremely well the observed power spectrum and its asymptotic scalings.

12 Primordial perturbations **ref**

According to our current leading paradigm, the *quantum* fluctuations on top of the classical inflationary background are the seeds of the structures that we see in our universe today. Hence we would like to quantize the inflationary model discussed in Sec. P.5.4. To this end, we start from the homogeneous background $\bar{g}_{\mu\nu}$ and $\bar{\phi}$, where $\bar{g}_{\mu\nu}$ is the FLRW spacetime whose scale factor $a(t)$ is given by the solution of the Friedmann equations (5.18), and $\bar{\phi}$ obeys the standard equations of motion in (5.17). Then we want to promote fluctuations to quantum operators

$$g_{\mu\nu}(t,\mathbf{x}) = \bar{g}_{\mu\nu}(t) + \hat{h}_{\mu\nu}(t,\mathbf{x}), \quad \phi(t,\mathbf{x}) = \bar{\phi}(t) + \hat{\varphi}(t,\mathbf{x}). \tag{12.1}$$

We will achieve this in two steps. First, we will quantize a free scalar field in curved spacetime, working mostly around de Sitter space. Second, we will see how to perturbatively quantize general relativity and hence predict the primordial power spectrum of curvature perturbations. Before jumping into the calculation, let's set some qualitative expectations.

12.1 The big picture

A key kinematical consequence of inflation is a shrinking comoving Hubble radius

$$\frac{d}{dt}(aH)^{-1} = \dot{r}_H < 0. \tag{12.2}$$

which is equivalent to having $w < -1/3$. As a consequence, a comoving perturbation k, which is sub-Hubble, is eventually stretched to super-Hubble scales by the accelerated expansion. On super-Hubble scales it remains constant until it re-enters the Hubble radius during decelerated expansion, in radiation or matter domination, providing initial conditions for all perturbations. An overview of this process is shown in Fig. 23.

The mechanism to achieve this in standard inflationary models is the dynamics of a scalar field ϕ, which is slowly rolling down its potential $V(\phi)$. How do quantum fluctuations affect this picture? As illustrated in Fig. 24, the inflaton value can be thought of as a clock that parametrizes the time to the end of inflation. Quantum fluctuations induce space-dependent variations of this clock compared to its averaga, $\bar{\phi}(\mathbf{x}) \to \bar{\phi} + \delta\phi(\mathbf{x})$. This then implies that locally inflation ends at different times $\delta t(\mathbf{x}) = -\frac{\delta\phi(\mathbf{x})}{\dot{\bar{\phi}}}$ – which in turn implies that different parts of space have expanded by different amounts. The curvature perturbation induced by inflationary fluctuations can be computed in flat gauge using:

$$\mathcal{R}(\mathbf{x}) \sim H dt(\mathbf{x}) \sim \frac{H}{\dot{\bar{\phi}}}\delta\phi(\mathbf{x}) \tag{12.3}$$

This quantity is evaluated at horizon exit $\mathbf{k} = aH$ for each mode $\delta\phi_{\mathbf{k}}$. Outside Hubble $\mathcal{R}$ mode is frozen. The fact that $\mathcal{R}$ is conserved outside the horizon is crucial for connecting early and late times in this calculation, because physics, especially inside the horizon, is quite uncertain at early times.

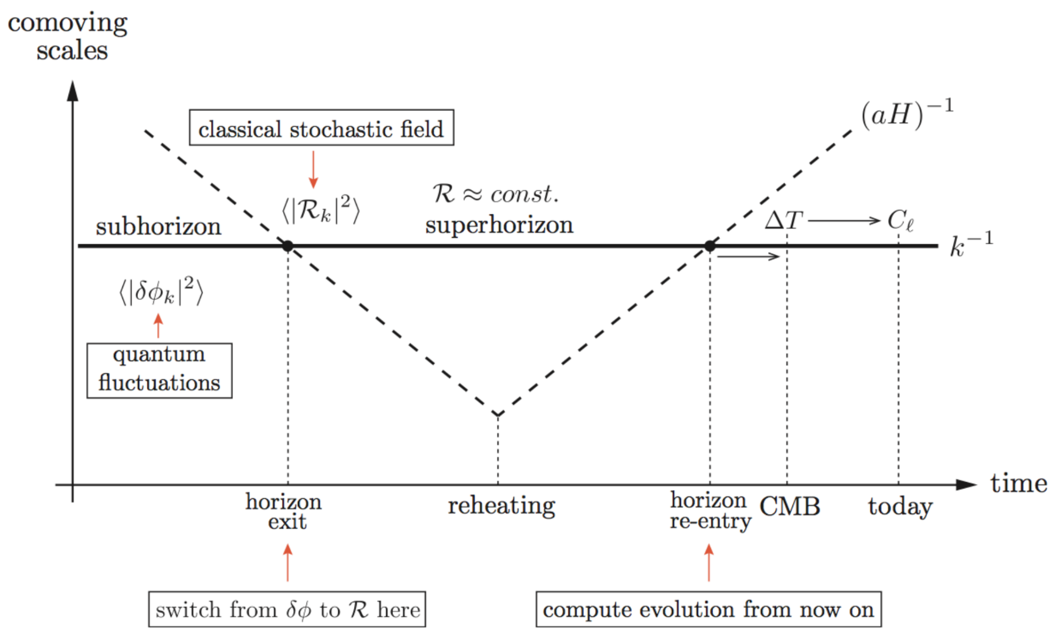


Figure 23: The evolution of perturbations during and after inflation. During inflation, the comoving Hubble radius shrinks while during the hot big bang it grows again. This means that a mode begins inside Hubble, exits the Hubble radius during inflation, and then re-enters the during the late Universe, during radiation or matter domination. The conservation of the curvature perturbation $\mathcal{R}$ is crucial in relating the mode at early times to late time observables. (Image credit: D. Baumann).

12.2 Inflaton fluctuations

A good starting point is the simple toy model of a minimally coupled scalar field ϕ in de Sitter spacetime. Consider the action

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right] , \tag{12.4}$$

where ϕ is *minimally coupled* to the metric, i.e. we simply started from the Minkowski theory and promoted $\eta_{\mu\nu}$ to $g_{\mu\nu}$ and d^4x to $d^4x\sqrt{-g}$. For later convenient we note now that the energy-momentum tensor for the theory above is

$$T_{\mu\nu} = \frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}} = \partial_\mu\phi\partial_\nu\phi + g_{\mu\nu}\left(-\frac{1}{2}\partial_\rho\phi\partial^\rho\phi - V\right) . \tag{12.5}$$

This takes the same form as the energy-momentum tensor of a perfect fluid, namely (1.34), under the following identifications

$$\rho = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi + V\,, \qquad p = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V\,, \qquad u_\mu = -\frac{\partial_\mu\phi}{\sqrt{-\partial_\rho\phi\partial^\rho\phi}} . \tag{12.6}$$

It turns out that we gravitational effects are subleading and we approximate the metric as just the homogeneous background $g_{\mu\nu} = \bar{g}_{\mu\nu}$. Now we expand this action around a classical

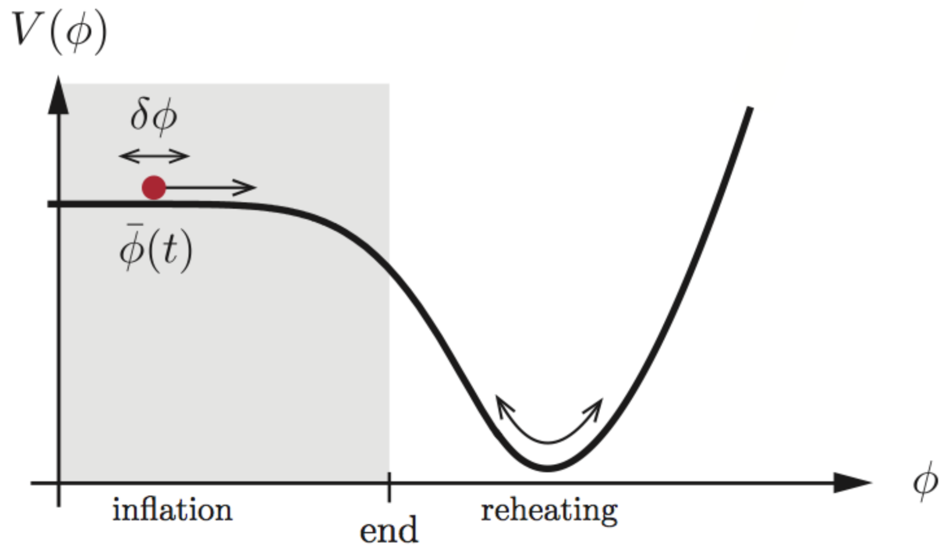


Figure 24: The inflation field rolling slowly down a potential. The field value can be thought of as a "clock" parametrizing the time until the end of inflation. Fluctuations in this field thus lead to fluctuation in the time throughout which the universe inflates. This locally varying expansion induces a primordial curvature perturbation $\mathcal{R}$, which represents the initial condition for the evolution of our universe. (Image credit: D. Baumann.)

solution $\bar{\phi}$, i.e. $\phi = \bar{\phi} + \varphi$. We can drop terms that do not depend on φ and the terms linear in φ will cancel by virtue of the background equations of motion for $\bar{\phi}$. Focussing on the terms quadratic in φ and neglecting higher orders we obtain

$$S = -\int \sqrt{-g}\frac{1}{2}\left[g^{\mu\nu}\partial_\mu\varphi\partial_\mu\varphi - V''\varphi^2\right] = \int d^3x dt\, a^3\frac{1}{2}\left(\dot{\varphi}^2 - \frac{1}{a^2}\partial_i\varphi\partial_i\varphi - \frac{1}{2}V''\varphi^2\right) . \quad (12.7)$$

For our final goal of computing the power spectrum of $\mathcal{R}$ we will only need to evaluate the dynamics of φ until Hubble crossing, i.e. for $k \gg aH$. In this regime the mass term V'' is smaller than the other terms

$$V''\varphi^2 \sim \eta_V H^2\varphi^2 \ll H^2\varphi^2 < \frac{k^2}{a^2}\varphi^2 ,$$

where η_V is the second potential slow-roll parameter, which is small during slow-roll inflation. Therefore we will neglect this term.

In Fourier space, this free theory reduces to an infinite sum over $\mathbf{k}$ of decoupled harmonic oscillators

$$S = \int \frac{d^3k}{(2\pi)^3} dt\, a^3\frac{1}{2}\left[\dot{\varphi}(\mathbf{k})\dot{\varphi}(-\mathbf{k}) - \frac{k^2}{a^2}\varphi(\mathbf{k})\varphi(-\mathbf{k})\right] , \quad (12.8)$$

where

$$\varphi(\mathbf{x}) = \int_{\mathbf{k}} \varphi(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}, \quad \varphi(\mathbf{k}) = \int_{\mathbf{x}} \varphi(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}} , \quad (12.9)$$

with

$$\int_{\mathbf{k}} \equiv \int \frac{d^3\mathbf{k}}{(2\pi)^3}\,, \qquad \int_{\mathbf{x}} \equiv \int d^3\mathbf{x}\,. \tag{12.10}$$

To quantize the theory, we promote φ to an operator (but we will omit the hat). As we do for the quantum harmonic oscillator, we write φ in terms of creation and annihilation operators

$$\varphi(\mathbf{k},t) = f_k(t)a_{\mathbf{k}} + f_k^*(t)a^\dagger_{-\mathbf{k}}\,, \tag{12.11}$$

which satisfy

$$[a_{\mathbf{k}}, a_{\mathbf{k}'}] = 0\,, \qquad [a_{\mathbf{k}}, a^\dagger_{\mathbf{k}'}] = (2\pi)^3\delta^3_D\left(\mathbf{k}-\mathbf{k}'\right)\,. \tag{12.12}$$

Notice that $\varphi^*(\mathbf{k}) = \varphi(-\mathbf{k})$, as required by the fact that $\varphi(\mathbf{x})$ is a real field. Here we have chosen to work in the Heisenberg picture, in which operators depend on time while the states do not, so $\varphi = \varphi(\mathbf{k},t)$ (but we'll omit the time argument when no ambiguity arises). All the time dependence of $\varphi(t,\mathbf{k})$ has been collected into $f_k(t)$ and $f_k^*(t)$, which are known as *mode functions*. They are determined by requiring that φ solves the equations of motion[95] derived from (12.8)

$$\ddot{f}_k + 3H\dot{f}_k + \frac{k^2}{a^2}f_k = 0\,. \tag{12.14}$$

Because of the isotropy of the background f_k depends only on the norm k of $\mathbf{k}$, as suggested by the notation. This equation becomes more familiar if we use conformal time ($' \equiv \partial_\tau$)

$$(af_k)'' + \left(k^2 - \frac{a''}{a}\right)(af_k) = 0\,, \tag{12.15}$$

where it looks like a harmonic oscillator with a time dependent mass $a''/a = 2/\tau^2$. The two linearly independent solutions are the complex conjugate of each other

$$f_k = \alpha(1+ik\tau)e^{-ik\tau} + \beta(1-ik\tau)e^{ik\tau}\,. \tag{12.16}$$

The integration constants α and β must be compatible with the canonical commutation relations. Let Π be the momentum conjugate[96] of φ,

$$\Pi(\mathbf{x}) \equiv \frac{\partial\mathcal{L}}{\partial\dot{\varphi}(\mathbf{x})} = \frac{\partial}{\partial\dot{\varphi}(\mathbf{x})}\left(\frac{1}{2}a^3\dot{\varphi}^2\right) = a^3(\tau)\dot{\varphi}(\mathbf{x}) = a^2(\tau)\partial_\tau\varphi(\mathbf{x})\,. \tag{12.17}$$

Then, canonical quantization imposes the constraint

$$[\hat{\varphi}_{\mathbf{k}}, \hat{\Pi}_{\mathbf{k}'}] = a^2\left(f_k f_k'^* - f_k^* f_k'\right)(2\pi)^3\delta^3\left(\mathbf{k}+\mathbf{k}'\right) \stackrel{!}{=} i(2\pi)^3\delta^3\left(\mathbf{k}+\mathbf{k}'\right)\,, \tag{12.18}$$

which applied to (12.16) in turn implies

$$|\alpha|^2 - |\beta|^2 = \frac{H^2}{2k^3}\,. \tag{12.19}$$

[95] Or equivalently the Heisenberg equation

$$\dot{\varphi}(\mathbf{k}) = i[H, \varphi(\mathbf{k})]\,. \tag{12.13}$$

where H is the Hamiltonian derived from the action (12.8).

[96] Note Π would be the same if we computed it using conformal time, i.e. $\Pi = \partial\mathcal{L}/\partial\varphi'$, where now $\mathcal{L} = a^2\varphi'^2/2$.

Initial conditions To completely determine the integration constants[97] α and β we notice that in the far past, i.e. for $k\tau \gg 1$, (12.15) reduces to the Klein-Gordon equation for the field (af_k), since $k^2 \gg a''/a$. In other words, the field $(a\varphi(\mathbf{k}))$ at early times lives effectively in Minkowski spacetime. In this limit we expect to recover the (Heisenberg picture) free scalar field that we learn about in introductory QFT,

$$\varphi(\mathbf{x},t) = \int \frac{d^3k_p}{(2\pi)^3} \frac{e^{i\mathbf{k}_p\cdot\mathbf{x}}}{\sqrt{2k_p}} \left[e^{-ik_pt} a_{\mathbf{k}_p} + e^{ik_pt} a^\dagger_{-\mathbf{k}_p}\right] \quad \text{(Minkowski)}\,, \tag{12.21}$$

where k_p is the physical wavenumber, related to the comoving one at some time by $k_p = k/a$. Recall that this choice of time dependence means that $\varphi(\mathbf{k}_p) \supset e^{+ik_pt} a^\dagger_{\mathbf{k}_p}$ *creates particles of positive energy*, as can be check from

$$\hat{H}\varphi(\mathbf{k})\left|0\right\rangle = [\hat{H}, \varphi(\mathbf{k})]\left|0\right\rangle + \varphi(\mathbf{k})\hat{H}\left|0\right\rangle = -i\dot{\varphi}(\mathbf{k})\left|0\right\rangle = +k\varphi(\mathbf{k})\left|0\right\rangle\,, \tag{12.22}$$

where the hat distinguishes the Hamiltonian $\hat{H}$ from the Hubble parameter H, and in the second step we used the Heisenberg equation and $\hat{H}\left|0\right\rangle = 0$. Comparing to (12.11), we can say that $f_k = e^{-ikt}$ are the Minkowski mode functions, the well known monochromatic solutions of the Klein-Gordon equation.
To find α and β we therefore match the solution (12.16) of the de Sitter equation of motion and its time derivative to the Minkowski mode functions, at some very early time τ_* where $k\tau_* \gg 1$,

$$af_k\Big|_{\tau=\tau_*} = \frac{e^{-ik_pt}}{\sqrt{2k_p}}\Big|_{t=t_*}\,, \qquad \partial_t(af_k)\Big|_{\tau=\tau_*} = \partial_t\left(\frac{e^{-ik_pt}}{\sqrt{2k_p}}\right)\Big|_{t=t_*}\,. \tag{12.23}$$

Solving the linear system for α and β one finds

$$\alpha = ie^{ik\tau_*(1+Ht_*)}\frac{H}{\sqrt{2k^3}}\left[1 + \frac{i}{k\tau_*} - \frac{1}{2(k\tau_*)^2}\right]\,, \tag{12.24}$$

$$\beta = ie^{-ik\tau_*(1-Ht_*)}\frac{H}{\sqrt{2k^3}}\frac{1}{2(k\tau_*)^2}\,. \tag{12.25}$$

If the matching is done in the infinite past, $k\tau_* \to -\infty$, this reduces simply to

$$\lim_{\tau\to-\infty}|\alpha| = \frac{H}{\sqrt{2k^3}}\,, \qquad \lim_{\tau\to-\infty}\beta = 0\,. \tag{12.26}$$

The normalization of α is fixed only up to an overall phase because one can always shift t in Minkowski and so the value of t_* is arbitrary. The dS mode functions that create positive-energy

[97] An alternative derivations goes as follows. Inverting (12.11) for $\hat{a}_\mathbf{k}$,

$$\hat{a}_\mathbf{k} = af_k^*\left(\frac{i\hat{\Pi}_\mathbf{k}}{a} - \frac{i\partial_\tau f_k^*}{f_k^*}\,a\hat{\varphi}_\mathbf{k}\right) \tag{12.20}$$

we find that $\hat{a}_\mathbf{k}$ annhilates the vacuum in the infinite past iff $-i\partial_\tau f_k^* \to f_k^* k$ as $\tau \to -\infty$. This implies that $f_k^* \sim e^{+ik\tau}$ asymptotically, which can be recognised as the mode function on Minkowski space for particles with positive energy $+k$. This requirement fixes $\beta = 0$.

particles in the infinite past therefore are

$$\boxed{f_k = \frac{H}{\sqrt{2k^3}}(1+ik\tau)e^{-ik\tau}} \quad \text{(dS mode functions)}\,. \tag{12.27}$$

The dS mode functions (12.27) oscillate for $k\tau \gg 1$, but become very different from their Minkowski counterpart when the comoving wavenumber becomes smaller than the comoving Hubble parameter

$$k \ll aH = \frac{1}{|\tau|} \quad \text{(Hubble crossing)}\,, \tag{12.28}$$

The swith between the two regimes takes place at *Hubble crossing*, when $k_{\text{H.c.}} = aH$, sometime also called horizon crossing. In physical length scales, this means the physical wavelength $\lambda_p = a/k$ is stretched by the expansion to become larger than the Hubble radius $1/H$. Since k and H are (approximately) constant, while $a = e^{Ht}$ grows with time, modes continuously cross the Hubble radius as time proceeds and become "super-Hubble" modes. Unlike "sub-Hubble" modes, $k \gg aH$, which oscillate, superHubble modes *freeze out* and asymptotes a constant

$$\lim_{\tau\to 0} f_k(\tau) = \frac{H}{1+\sqrt{2k^3}}\left(\frac{1}{2}k^2\tau^2 - \frac{1}{3}ik^3\tau^3 + \dots\right)\,. \tag{12.29}$$

We quoted the subleading terms to highlight the fact that the imaginary part of f, i.e. the difference between f and f^*, is suppressed by $(k\tau)^3$.

More rigorous ways to derive (12.27) include Hamiltonian minimization, i.e. choosing as vacuum the lowest energy state in the asymptotic past and matching to the uniquely defined Euclidean vacuum of the Wick rotated Euclidean field theory. Now that we have related φ to creation and annihilation operators, we can specify the "vacuum state" $|0\rangle$ by the usual condition defining

$$\text{Fock vacuum:} \qquad a_{\mathbf{k}}\,|0\rangle = 0\,, \tag{12.30}$$

for all $\mathbf{k}$. In this context, the state $|0\rangle$ is called the *Bunch-Davies state*[98] [11]. Excited states are then obtained by acting with creation operators on this vacuum.

Expectation values and the power spectrum What observables can we compute for this theory? As familiar from Quantum Mechanics, observables are given by the expectation value of operators. In cosmology, we have observational access only to these expectation values in the infinite future $k\tau \to 0$. As we have seen, when stated in terms of curvature perturbations$\mathcal{R}$, in this limit correlators become approximately constant and so we will only be interested in the expectation value of product of correlators at equal time, or simply cosmological correlators for short

$$\lim_{\tau\to 0}\langle \mathcal{O}(\mathbf{k}_1,\tau)\dots\mathcal{O}(\mathbf{k}_n,\tau)\rangle\,, \tag{12.31}$$

[98] It also coincides with the Hartle-Hawking state [34] encountered when solving the Wheeler-deWitt equation [21].

Box 12.1 Free theories and Gaussianity All free theories can be understood by analogy with the most famous free theory, the quantum harmonic oscillator. In quantum mechanics you learn that the probability of finding a particle at position x is a Gaussian distribution

$$\text{Prob}(x) \sim |\psi(x)|^2 \propto e^{-x^2/(2\sigma^2)}\,, \tag{12.32}$$

where $\psi(x)$ is the position space wavefunction and $\sigma^2 = \langle x^2\rangle$ describes the width of the Gaussian and is fixed by the parameters of the theory ($\sigma^2 = (2m\omega)^{-1}$). It is clear by parity that

$$\langle x^{2n+1}\rangle \propto \int_{-\infty}^{\infty} dx\,|\psi(x)|^2 x^{2n+1} \sim \int dx\, e^{-x^2/(2\sigma^2)} x^{2n+1} = 0\,. \tag{12.33}$$

While with repeated integration by parts we can always rewrite $\langle x^{2n}\rangle$ in terms of $\langle x^2\rangle^n$. Because of this, the expression free theory and Gaussian theory are often used interchangeably. In the next Sec. we will study interacting theories, where we will compute *non-Gaussianities*, i.e. deviations from a Gaussian wavefunction. This discussion applies to quantum*field* theory (as opposed to quantum mechanics) by thinking of $\phi(\mathbf{k})$ as an infinite collection of decoupled harmonic oscillators, as depicted in Fig. 25

for some local operators $\mathcal{O}$. Because we are studying a *free* theory, all information is contained in the two-point correlators of φ and its conjugate momentum π. All odd-point correlators vanish by the symmetry $\varphi \to -\varphi$ and all higher even-point correlators can be reduced to the two-point one using Wick's theorem.
Let's compute the two-point correlation function of φ in the Bunch-Davies state:

$$\lim_{\tau\to 0}\langle \varphi(\mathbf{k},\tau)\varphi(\mathbf{k}',\tau)\rangle = |f_k(,\tau)|^2 \langle 0|\, a_{\mathbf{k}} a^\dagger_{-\mathbf{k}'}\,|0\rangle = (2\pi)^3\delta^3_D\left(\mathbf{k}+\mathbf{k}'\right)P(k,\tau) \tag{12.34}$$

with

$$P(k,\tau) = \frac{H^2}{2k^3}\left[1+(k\tau)^2\right]\,. \tag{12.35}$$

In the late time limit, or equivalently in the limit in which we are looking at a perturbation well outside the Hubble radius $k|\tau| \ll 1$, this reduces to the scale-invariant power spectrum

$$\lim_{\tau\to 0} P(k,\tau) \equiv P(k) = \frac{H^2}{2k^3}\,. \tag{12.36}$$

Here, we have introduced the late-time *power spectrum* $P(k)$, which is just the two-point correlator stripped of the Dirac delta and its accompanying factor of $(2\pi)^3$. A few comments are in order:

- The Dirac delta reminds us that momentum is conserved as a consequence of the homogeneity of the background. Pictorially we can imagin that perturbations in this state must exist in pairs of opposite wavenumber $\mathbf{k}$ and $-\mathbf{k}$.

- $P(k)$ does not depend on the direction of $\mathbf{k}$ as consequence of the isotropy of the background.

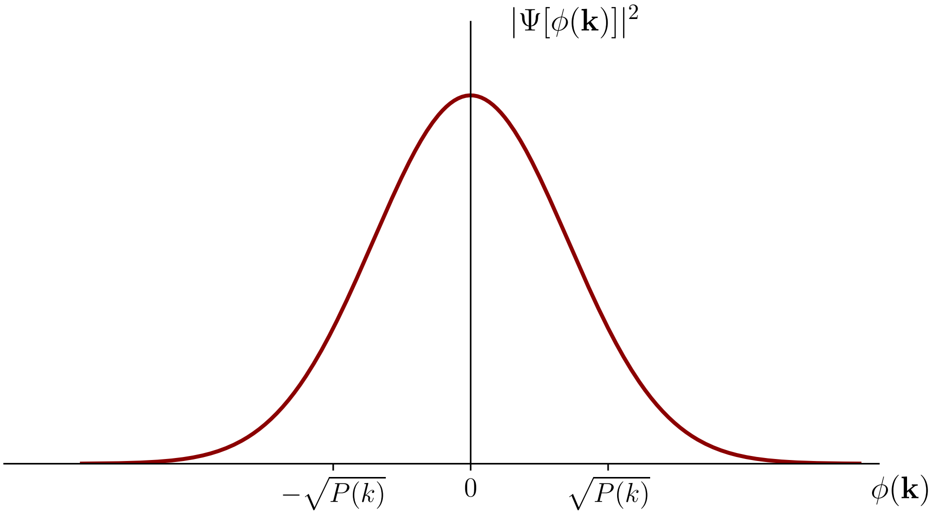


Figure 25: The figure show that the probability distribution function of a free field theory, which is proportional to the norm square of the wavefunction, is a multivariate Gaussian in the infinitely many decoupled harmonic oscillators $\phi(\mathbf{k})$, each with variance $P(k)$.

- As we will see shortly, the fact that the power spectrum asymptotes some (non-vanishing) constant value as $\tau \to 0$ is related to the absence of a mass.
- The k-dependence $P \propto k^{-3}$ is the one corresponding to *scale invariance*, as defined in (4.26). To see this, we can Fourier transform to the real-space *correlation function*,[99]

$$\langle \varphi(\mathbf{x})\varphi(\mathbf{0})\rangle = \int_{\mathbf{k}\mathbf{k}'} \langle \varphi(\mathbf{k})\varphi(\mathbf{k}')\rangle \sim H^2 \,, \tag{12.38}$$

and notice that the correlation does not depend on distance. In particular, it doesn't change if we rescale $\mathbf{x} \to \lambda \mathbf{x}$.

It is worth comparing the power spectrum in dS, (12.36), with that in Minkowski

$$\langle \varphi(\mathbf{k})\varphi(\mathbf{k}')\rangle = (2\pi)^3 \delta_D^3 \left(\mathbf{k}+\mathbf{k}'\right) \frac{1}{2k} \quad \text{(Minkowski)}\,. \tag{12.39}$$

and the associated real-space correlation function

$$\langle \varphi(\mathbf{x})\varphi(\mathbf{0})\rangle = \int_{\mathbf{k}\mathbf{k}'} \langle \varphi(\mathbf{k})\varphi(\mathbf{k}')\rangle \sim \frac{1}{x^2} \quad \text{(Minkowksi)}\,. \tag{12.40}$$

[99] Actually the dS correlation function at separated points, $\mathbf{x} \neq 0$, is IR divergent. The physical reason is that dS is eternal. This divergence can be regularized either with a small tilt of the power spectrum $k^{-(3+\delta)}$, with $0 < \delta \ll 1$ or with an IR cutoff $k_{\rm IR}$ of the integral ($\tilde{k} = xk$ is dimensionless)

$$\langle \varphi(\mathbf{x})\varphi(\mathbf{0})\rangle = \frac{H^2}{(2\pi)^2} \int_{xk_{\rm IR}}^{\infty} d\tilde{k}\, \frac{\sin \tilde{k}}{\tilde{k}^2} \xrightarrow{xk_{\rm IR}\to 0} H^2 \left[\gamma_E - 1 + \log(xk_{\rm IR}) + \mathcal{O}((xk_{\rm IR})^2)\right] \,. \tag{12.37}$$

.

So in dS the correlation function is independent of the distance, while in Minkowski it decays as $1/x^2$, as expected for a massless particle. For completeness, we quote the correlators involving the momentum conjugate:

$$\lim_{\tau\to 0}\langle\varphi(\mathbf{k})\Pi(\mathbf{k}')\rangle = (2\pi)^3\delta_D^3\left(\mathbf{k}+\mathbf{k}'\right)\frac{1}{2k\tau}\,, \tag{12.41}$$

$$\lim_{\tau\to 0}\langle\varphi(\mathbf{k})\Pi(\mathbf{k}')\rangle = \lim_{\tau\to 0}\langle\pi(\mathbf{k})\varphi(\mathbf{k}')\rangle\,, \tag{12.42}$$

$$\lim_{\tau\to 0}\langle\pi(\mathbf{k})\Pi(\mathbf{k}')\rangle = (2\pi)^3\delta_D^3\left(\mathbf{k}+\mathbf{k}'\right)\frac{k}{2H^2\tau^2}\,. \tag{12.43}$$

12.3 Massive scalar in de Sitter*

It is interesting to ask what changes if the scalar field has a mass[100] m,

$$S = -\int\sqrt{-g}\frac{1}{2}\left[\partial_\mu\varphi\partial^\mu\varphi + m^2\varphi^2\right]\,. \tag{12.44}$$

You will quantize this theory in Example Sheet 1. The relation to creation and annihilation is the same

$$\varphi(\mathbf{k},t) = f_k(t)a_{\mathbf{k}} + f_k^*(t)a^\dagger_{-\mathbf{k}}\,, \tag{12.45}$$

but now the mode functions are modified to

$$f_k(\tau) = i\frac{\sqrt{\pi}H}{2}(-\tau)^{3/2}H_\nu^{(1)}(-k\tau)\,,\qquad \nu = \sqrt{\frac{9}{4}-\frac{m^2}{H^2}}\,, \tag{12.46}$$

where $H^{(1)}$ is the Hankel function of the first kind and we are free to multiply this by any complex phase $e^{i\eta}$. Hankel functions are solutions of the Bessel's differential equation and are linear combination of Bessel functions and in general they cannot be expressed in terms of elementary functions. However when ν is half integer, $H_\nu^{(1,2)}(x)$ reduces to an oscillating phase $e^{\pm ix}$ times $x^{-\nu}$ times a polynomial in x. An interesting case is $m = \sqrt{2}H$ so that $\nu = 1/2$ and

$$f_k(\tau) = \frac{H}{\sqrt{2k}}\tau e^{-ik\tau}\,. \tag{12.47}$$

This value of the mass corresponds to a conformally coupled scalar, namely a scalar non-minimally coupled to gravity with a $\varphi^2 R$ term in such a way that the full theory is invariant under a Weyl rescaling of the metric $g_{\mu\nu}\to\Omega(x)g_{\mu\nu}$. Since the de Sitter metric is proportional to that of Minkowski, where the Ricci scalar vanishes, this theory is equivalent to a massless scalar in Minkowski. This explains why the conformally coupled mode function in (12.47) is just the rescaled Minkowski mode function in (12.23).

[100] The insightful reader might wonder about the true meaning of mass in dS. Indeed, in Minkowski, mass can be defined very generally as one of the two parameters (eigenvalues of the Casimir operators) classifying the unitary irreps of the Poincaré group (the other parameter being spin for $m^2 > 0$ or helicity for $m = 0$). In dS this is not possible (see [29] for an in depth discussion). Here by mass we simply mean the parameter m appearing in (12.44). In particular we assume that the scalar field is minimally coupled to gravity, e.g. there is no coupling to the Ricci scalar of the form $\xi R\phi$, and scale factor appears only because of $\sqrt{-g}$ and $g^{\mu\nu}$.

Back to general ν. For $\tau \to 0$ this becomes

$$f_k(\tau \to 0) = H\tau^{3/2} \left[\frac{\sqrt{\pi}(-k\tau)^{\nu}}{2^{1+\nu}\Gamma[\nu+1]}(1 + i\cot(\pi\nu)) - \frac{i(-k\tau)^{-\nu}}{\sqrt{\pi}2^{1-\nu}\Gamma[\nu]} \right] + \dots . \tag{12.48}$$

Now it is useful to distinguish two cases. The first case is when the mass square is small or negative, $m^2 < 9H^2/4$. Then ν is real and positive. In this case, the first term in brackets approaches zero faster than the second and can be neglected. So the power spectrum now becomes

$$P(k) = |f_k|^2 = \frac{H^2}{\pi 2^{2(\nu-1)}\Gamma(\nu)^2} \frac{(-k\tau)^{3-2\nu}}{k^3} \quad (\text{for } m^2 < \frac{9}{4}H^2) . \tag{12.49}$$

Because of the mass, the power spectrum is not scale invariant anymore, $P \propto k^{-2\nu}$. Also, P has acquired a time dependence. For positive $m^2 > 0$, one finds $3-2\nu > 0$ and the power spectrum decays with time and vanish at future infinity. This is to be expected because the quadratic potential pushes the field towards $\varphi = 0$. For negative m^2 we would expect an instability and indeed the power spectrum grows with time and diverges at future infinity.
The second case is when the mass square is large and positive, $m^2 > 9H^2/4$, then ν becomes complex and the two terms in the brackets of (12.48) are of the same order. The power spectrum oscillates while decaying as τ^3. In cosmology, we are mostly interested in massless or almost massless fields, which do not create large instability and whose perturbations survive long enough to be observable at late times.

12.4 Particle creation*

For QFT in Minkowski, we can think of excitations generated by the creation operators as particles. However, in curved spacetime particle and particle number are more subtle concepts. Let's see this in detail. We found field excitations in dS oscillate as $f_k \sim e^{-ik\tau}$, where the comoving wavenumber k is related to the Minkowski energy by

$$E = \sqrt{k_i k_j g^{ij}} = \frac{k}{a} . \tag{12.50}$$

Let us now Taylor expand the time-dependent phase of f_k in time around some time t_*:

$$\begin{aligned} -ik\tau &= i\frac{k}{aH} = i\frac{k}{a_*}\frac{a_*}{aH} \\ &= i\frac{E}{H}e^{-H(t-t_*)} \simeq iE\left[\frac{1}{H} - (t-t_*) + \frac{1}{2}H(t-t_*)^2 + \dots\right] . \end{aligned} \tag{12.51}$$

The first term in brackets is an irrelevant phase. The second term is precisely the time dependence of particles in Minkowski, namely e^{-iEt}. So field excitations in dS have a chance to look like particles only for a time interval $(t - t_*) \ll 1/H$, during which we can neglect the higher order terms in brackets. Moreover, we must demand that during this interval, the wavefunction oscillates many times, so $E(t - t_*) \gg 1$. Using again (12.51) this requires $E/H = -k\tau \gg 1$. When this condition is not satisfied, the energy and momentum of the state are redshifted by the expansion before a single oscillation of the wavefunction.

Even when particles can be defined, the expansion of the universe can create particles (unless there is a conserved quantum number) at a rate controlled again by H. This is to be expected as the expansion of the universe breaks time translations and so energy is not conserved. Let's see this in more detail.
In the previous section, we found the mode functions by demanding that φ creates positive-energy particles at $k\tau \to -\infty$. Let's instead require that φ creates positive-energy particles at some finite τ_*, still satisfying $|k\tau_*| \gg 1$. By matching to the Minkowski vacuum at τ_*, we find α and β as given in (12.24). The quantized field then takes the form

$$\varphi(\mathbf{k}) = g_k b_{\mathbf{k}} + g_k^* b_{-\mathbf{k}}^\dagger \tag{12.52}$$

with[101]

$$\begin{aligned} g_k &= \alpha f_k(\tau) + \beta f_k^*(\tau) \\ &= \frac{H}{\sqrt{2k^3}}\left[1 + \frac{i}{k\tau_*} - \frac{1}{2(k\tau_*)^2}\right] f_k(\tau) + e^{-2ik\tau_*}\frac{H}{\sqrt{2k^3}}\frac{1}{2(k\tau_*)^2} f_k^*(\tau)\,, \end{aligned} \tag{12.53}$$

and $\{b_{\mathbf{k}}, b_{\mathbf{k}}^\dagger\}$ a new set of creation and annihilation operators, which define a new vacuum state by $b_{\mathbf{k}}\left|\tilde{0}\right\rangle = 0$. By matching the two expressions for $\varphi(\mathbf{k})$, (12.11) and (12.52), we see that the two sets of ladder operators are related,

$$a_{\mathbf{k}} = \frac{\sqrt{2k^3}}{H}\left(\alpha b_{\mathbf{k}} + \beta^* b_{-\mathbf{k}}^\dagger\right)\,, \qquad a_{\mathbf{k}}^\dagger = \frac{\sqrt{2k^3}}{H}\left(\beta b_{-\mathbf{k}} + \alpha^* b_{\mathbf{k}}^\dagger\right)\,, \tag{12.54}$$

This relation is called a *Bogoliubov transformation.* It can be inverted to give

$$b_{\mathbf{k}} = \frac{\sqrt{2k^3}}{H}\left(\alpha^* a_{\mathbf{k}} + \beta^* a_{-\mathbf{k}}^\dagger\right)\,, \qquad b_{\mathbf{k}}^\dagger = \frac{\sqrt{2k^3}}{H}\left(\beta a_{-\mathbf{k}} + \alpha a_{\mathbf{k}}^\dagger\right)\,, \tag{12.55}$$

Now we want to ask what a detector that measures $b_k^\dagger$ excitations would measure in the Bunch Davies vacuum. To this end, we define the "b-particle" number operator

$$N_b(\mathbf{k}) = b_{\mathbf{k}}^\dagger b_{\mathbf{k}}\,. \tag{12.56}$$

As expected, this operator has a vanishing expectation value in the $\left|\tilde{0}\right\rangle$ state. But if we compute its expectation value in the Bunch-Davies vacuum $|0\rangle$ we find

$$\langle 0|\, N_b(\mathbf{k})\,|0\rangle = \frac{2k^3}{H^2}|\beta_k|^2(2\pi)^3\delta_D^3(\mathbf{0}) = \frac{1}{4(k\tau)^4}(2\pi)^3\delta_D^3(\mathbf{0})\,. \tag{12.57}$$

The singular factor $\delta_D^3(\mathbf{0})$ is a reminder that we are working with an infinite volume

$$(2\pi)^3\delta_D^3(\mathbf{0}) = \lim_{V\to\infty}\int_V d^3x e^{-i\mathbf{0}\cdot\mathbf{x}} = \lim_{V\to\infty} V\,. \tag{12.58}$$

It is therefore wise to compute the number density of particles, $n_b(\mathbf{k}) \equiv N_b(\mathbf{k})/V$, instead of the total number $N_b(\mathbf{k})$. We find

$$\langle 0|\, n_b(\mathbf{k})\,|0\rangle = \frac{1}{4(k\tau)^4} \neq 0\,. \tag{12.59}$$

[101] Again, we used the arbitrariness in t_* to multiply g_k by a convenient phase.

In words, the Bunch-Davies "vacuum" state is found to contain some b-particles, defined with respect to the "vacuum" state at some large but finite $|\tau_*|$. As we take $\tau_* \to -\infty$, $|\tilde{0}\rangle$ approaches $|0\rangle$ and indeed the number density of particles vanishes as expected. We can say that the expansion of the universe creates particles. These particles are always created in pairs of opposite wavenumber, to conserve momentum, $n(-\mathbf{k}) = n(\mathbf{k})$.

12.5 Gauge invariant variables: curvature perturbations

Instead of choosing a specific set of coordinates one can work with gauge invariant variables[102]. This is not particularly convenient during the calculation, but it is a useful way to present the final result.
The idea is to find combinations of variables whose gauge transformations cancel each other. There are infinitely many options. For example, it is easy to see from (8.42) that the following combinations are gauge invariant at linear order

$$\begin{aligned} &\gamma_{ij} \ , && \Phi_i = N_i^{(V)} - \frac{a^2}{2}\dot{C}_i \ , \\ &-2\Psi = A + Ha^2\left(2\psi - \dot{B}\right) \ , && -2\Phi = h_{00} - \partial_t\left(a^2\left(2\psi - \dot{B}\right)\right) \ . \end{aligned} \tag{12.60}$$

These are called the *Bardeen variables* and were introduced in [7]. In the inflationary literature it is more common to use alternative gauge invariant variables.

Curvature perturbations Recall that we defined the combination in (P.5.4). Since we are in the presence of a scalar field rather than a fluid, what should we use for δu. To answer this we look back at (12.6) and recognize that the scalar part of the fluid velocity is

$$u_i = \partial_i \delta u = -\partial_i \phi / \sqrt{-\partial_\rho \phi \partial^\rho \phi} \, .$$

Expanding to linear order around the inflaton background this gives us $\delta u = -\varphi/\dot{\bar{\phi}}$ and hence

$$\mathcal{R} \equiv \frac{A}{2} - \frac{H}{\dot{\bar{\phi}}}\varphi \, . \tag{12.61}$$

We also recall that curvature perturbations on comoving hypersurfaces $\mathcal{R}$ owe their name to the fact that that in comoving gauge we have $\delta u \propto \varphi = 0$ and so $\mathcal{R} = A/2$. Then $\mathcal{R}$ appears in the metric as

$$g_{ij} = a^2 dx^i \delta_{ij} dx^j (1 + 2\mathcal{R}) \quad \text{(comoving gauge)} \, . \tag{12.62}$$

Then the Ricci scalar of a spatial hypersurface is ${}^{(3)}R = -4\partial_i^2 \mathcal{R}$ and so $\mathcal{R}$ generates a position dependent curvature. Nomen est omen.

[102] By "gauge invariant" here we have mean something distinct and weaker than what is meant in discussions of quantum gravity. Here a variable is "gauge invariant" if it does not transform at linear or higher order in perturbation theory under a change of coordinates *after the new coordinate is identified with the old one* (see Sec. **??**). When discussing quantum gravity, "gauge invariant" has a stronger meaning and requires an observable to not transform at all as one changes coordinates. In this stronger sense a local scalar field operator is not gauge invariant, since $\phi(x) \to \phi(x - \epsilon)$. Hence there is no contradiction between our discussion here and the often stated fact that there are no local and gauge-invariant observables in quantum gravity.

Since $\mathcal{R}$ is gauge invariant, we can and will compute it in any gauge we want. In particular, in flat gauge $A = 0$, and so

$$\mathcal{R} = -\frac{H}{\dot{\bar{\phi}}}\varphi \quad \text{(flat gauge)}\,. \tag{12.63}$$

This relation teaches us something interesting. Notice that $H/\dot{\bar{\phi}}$ is a time dependent function. Its time derivative may be slow-roll suppressed but it is not zero. This implies that $\mathcal{R}$ and φ cannot be both constant in time. Since we proved in great generality that $\mathcal{R}$ is constant on superHubble scales, we conclude that φ must evolve during inflation. This is another important reason why the predictions of the early universe should always be computed in terms of superHubble correlators of $\mathcal{R}$, which is gauge invariant and conserved, rather than those of φ, which evolves with time and changes from gauge to gauge.

Beyond linear order As a side remark, here we briefly mention that happens beyond linear order. It would be nice to define a variable that is gauge invariant not just to linear order. Indeed, a tree-level n-point function of $\mathcal{R}$ is gauge invariant iff $\mathcal{R}$ is gauge invariant to at least order $n-1$. With a bit of work we could find a second-order version of (12.63), but this involves a lot of algebra as we have to recompute all the gauge transformations to second order. Instead, we'll try to be clever. We'll define the gauge-invariant variable $\mathcal{R}$ to be the quantity that in comoving gauge appears in the metric as[103]

$$g_{ij} = a^2 e^{2\mathcal{R}}\delta_{ij} \quad \text{(comoving gauge)}\,. \tag{12.64}$$

In other gauges, $\mathcal{R}$ is given by its value in comoving gauge plus all the terms induced by the gauge transformation to the required order. This agrees with the previous definition (12.61) at linear, because $\varphi = 0$ and $e^{2\mathcal{R}} = 1 + A$.

The non-linear definition of $\mathcal{R}$ lends itself to a nice interpretation, which is depicted in Fig 26. In particular, we can think of the spatial variation of $\mathcal{R}(\mathbf{x})$ as modulating the local amount of expansion compared to the expansion $a(t)$ prescribed by the background. For example, in de Sitter a point $\mathbf{x}$ experience a local amount of expansion corresponding to a number of efoldings

$$N_e = \log a(t) e^{\mathcal{R}(\mathbf{x})} = Ht + \mathcal{R}(\mathbf{x})\,. \tag{12.65}$$

This means that the number of efoldings of expansion in a perturbed universe depends on the point, $N_e = N_e(\mathbf{x}, t)$. Then one can define perturbations to N_e as usual by subtracting the average $\delta N = N_e - \bar{N}_e$, where $\bar{N}_e$ is the spatial average of N_e. In an appropriate gauge, $\mathcal{R}$ coincides with δN to all orders. This is the starting point of the so-called δN formalism that describes the evolution of superHubble perturbations of $\mathcal{R}$ in an expansion in spatial derivatives [57,58].
In summary, gauge-invariant curvature perturbations $\mathcal{R}$ are linearly related to scalar field perturbations φ in flat gauge by the simple rescaling in (12.63). We will use this to express our results for the correlators of φ in terms of the correlators of $\mathcal{R}$, which is conserved on super-Hubble scales.

[103] This assumes we set tensors to zero, $\gamma_{ij} = 0$. Later one we will see that γ_{ij} can be included by the substitution $\delta_{ij} \to \exp(\gamma_{ij})$.

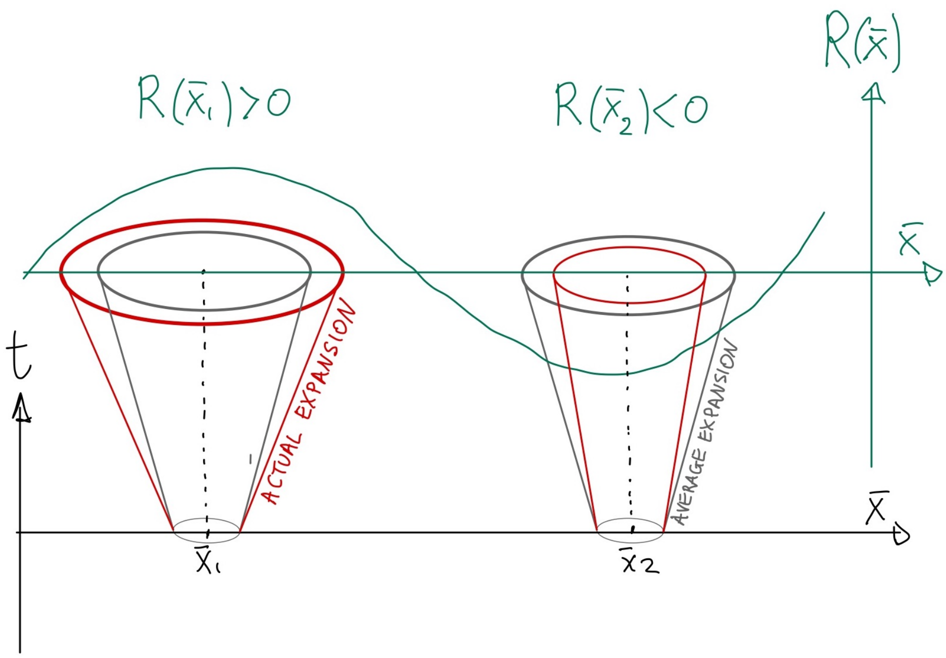


Figure 26: The figure shows the intuitive meaning of $\mathcal{R}(\vec{x})$ perturbations in comoving gauge. The unperturbed homogeneous universe expands in time (along the vertical axis) by the same amount at every point (the "average expansion" is represented by gray lines). But in the perturbed universe, points with different values of $\mathcal{R}(\vec{x})$ experience a larger (smaller) amount of expansion if $\mathcal{R}(x) > 0$ ($\mathcal{R}(x) < 0$) (the "actual expansion" is represented by red lines), as indicated in (12.65).

12.6 The primordial power spectrum

To conclude our discussion we would like to write down the correlators computed from $P(X,\phi)$ in terms of the gauge invariant variable $\mathcal{R}$. Using the power spectrum for φ in dS, (12.36), and the relation (12.63), we can compute the power spectrum of $\mathcal{R}$:

$$P_{\mathcal{R}}(k) = \frac{H^2}{\dot{\phi}^2} P_{\varphi}(k) = \frac{1}{4\epsilon}\left(\frac{H}{M_{\rm Pl}}\right)^2 \frac{1}{k^3}\,. \tag{12.66}$$

This an important result. It relates the amplitude of quantum fluctuations in $\mathcal{R}$, which in comoving gauge is a fluctuation of the metric, to the initial conditions for the distribution of all species in our Universe. The amplitude is fixed by the ratio $H^2/(\epsilon M_{\rm Pl}^2)$. Unfortunately we don't know the value of ϵ and so measurements of $P_{\mathcal{R}}$ cannot be used along to fix the scale H at which inflation took place.

The CMB tightly constrains the overall *amplitude* of the primordial curvature power spectrum. Observational analyses parameterise the dimensionless spectrum as

$$\Delta^2_{\mathcal{R}}(k) = A_s \left(\frac{k}{k_\star}\right)^{n_s-1}, \tag{12.67}$$

where A_s is the amplitude at the pivot scale $k_\star = 0.05\ \text{Mpc}^{-1}$. From the temperature and polarisation anisotropies measured by the Planck satellite, combined with large-scale structure and CMB lensing data, one finds

$$A_s = (2.10 \pm 0.03)\times 10^{-9} \qquad (68\%\ \text{CL})\,. \tag{12.68}$$

This $\mathcal{O}(10^{-9})$ amplitude is a key piece of evidence for inflation: it corresponds to quantum fluctuations of the inflaton field stretched to super-Hubble scales, and its smallness implies that inflation occurred at an energy scale far below the Planck scale. In fact, using (**??**) together with the slow–roll approximation and CMB constraints on r, one infers a characteristic Hubble scale $H \lesssim 6\times 10^{13}$ GeV during the observable window of inflation. The measured value of A_s therefore sets a non-trivial bound on any inflationary model, fixing the combination H^2/ϵ at the time the pivot mode exited the Hubble radius.

While the amplitude is measured with excellent precision, it is important to emphasise that it is only constrained over the limited range of scales probed by the CMB, roughly $k \sim 10^{-4}$–$0.2\ \text{Mpc}^{-1}$. On significantly smaller scales the primordial spectrum is not directly observable with current CMB experiments. Any extension of A_s beyond this window requires either an extrapolation of the nearly scale-invariant power law, or the use of complementary probes. Ultraviolet scales (such as those relevant for primordial black hole production or spectral distortions) are subject to comparatively weak bounds, many orders of magnitude weaker than those from the primary CMB. Infrared scales much larger than the current horizon are likewise unconstrained except by theoretical priors. Consequently, although A_s is known at the pivot scale with percent-level accuracy, the primordial amplitude on most cosmological scales remains effectively unknown, and significant deviations from a pure power law outside the CMB window remain compatible with current data.

12.7 The spectral tilt

At face value, the expression above seem to suggest an exactly scale invariant power spectrum, but this is incorrect. To understand why we have to think back at how we performed our calculation. First we computed the power spectrum of φ and then we converted it to that of $\mathcal{R}$. As we noticed, the conversion factors $H/\dot{\phi}$ is time dependent, so at what time are we supposed to evaluate it? Since we know that $\mathcal{R}$ becomes constant on super-Hubble scales, but not φ, the best time to evaluate it is at Hubble crossing, i.e. when $k = aH$ for each Fourier mode. So, for the dimensionless power spectrum $\Delta^2_{\mathcal{R}}(k)$ we should write

$$\Delta^2_{\mathcal{R}}(k) = \frac{k^3 P_{\mathcal{R}}(k)}{2\pi^2} = \frac{1}{8\pi^2}\left(\frac{H^2}{\epsilon M^2_{\mathrm{Pl}}}\right)_{aH=k} . \tag{12.69}$$

Because different modes cross Hubble at different times, the time-dependent coefficients in $\Delta^2_{\mathcal{R}}$ are in effect scale dependent. This introduces a small deviation from exact scale invariance that we call the *primordial spectral tilt.*
Exact scale invariance means that $\Delta^2_{\mathcal{R}}(k)$ is independent of k. Observationally, current CMB data are accurately described by a simple power law

$$\Delta^2_{\mathcal{R}}(k) = A_s \left(\frac{k}{k_\star}\right)^{n_s - 1} , \tag{12.70}$$

where A_s is the amplitude at the pivot scale $k_\star$, and n_s is the *scalar spectral index.* The spectral tilt is then defined as

$$n_s - 1 \equiv \left.\frac{\mathrm{d}\ln\Delta^2_{\mathcal{R}}(k)}{\mathrm{d}\ln k}\right|_{k=k_\star} . \tag{12.71}$$

For an exactly scale invariant spectrum $n_s = 1$, while $n_s < 1$ ($n_s > 1$) corresponds to a *red* (*blue*) spectrum with less (more) power on small scales.

The k dependence of $\Delta^2_{\mathcal{R}}$ arises because different k modes exit the Hubble radius at slightly different times, when H and ϵ have slightly different values. To make this dependence explicit, notice that for each mode

$$k = aH \quad \Rightarrow \quad \ln k = \ln a + \ln H . \tag{12.72}$$

Differentiating with respect to cosmic time we obtain

$$\frac{\mathrm{d}\ln k}{\mathrm{d}t} = H + \frac{\dot{H}}{H} = H(1-\epsilon) \simeq H , \tag{12.73}$$

From (12.69) we obtain

$$\frac{\mathrm{d}\ln\Delta^2_{\mathcal{R}}}{\mathrm{d}t} = 2\frac{\dot{H}}{H} - \frac{\dot{\epsilon}}{\epsilon} = -2\epsilon H - \frac{\dot{\epsilon}}{\epsilon} . \tag{12.74}$$

Combining this with the definition (12.71) and using $\mathrm{d}\ln k \simeq H\,\mathrm{d}t$ we find

$$n_s - 1 = \frac{\mathrm{d}\ln\Delta^2_{\mathcal{R}}/\mathrm{d}t}{\mathrm{d}\ln k/\mathrm{d}t} \simeq -2\epsilon - \frac{\dot{\epsilon}}{\epsilon H} = -2\epsilon - \eta . \tag{12.75}$$

where we used the second Hubble slow–roll parameter η. Both ϵ and η should be evaluated at Hubble crossing $k = aH$ for the mode $k = k_\star$.

Relation to the inflaton potential. The slow–roll parameters can be related to the shape of the inflaton potential $V(\phi)$ by using the background slow–roll equations

$$3H\dot{\bar{\phi}} \simeq -V_{,\phi}\,, \qquad 3M_{\rm Pl}^2 H^2 \simeq V(\bar{\phi})\,. \tag{12.76}$$

It is convenient to define the potential slow–roll parameters

$$\epsilon_V \equiv \frac{M_{\rm Pl}^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2, \qquad \eta_V \equiv M_{\rm Pl}^2 \frac{V_{,\phi\phi}}{V}\,. \tag{12.77}$$

Under the slow–roll approximation one finds $\epsilon \simeq \epsilon_V$ and, to leading order,

$$n_s - 1 \simeq -6\epsilon_V + 2\eta_V\,. \tag{12.78}$$

We see that the tilt is controlled directly by the slope and curvature of the potential. Very flat, slowly varying potentials (small ϵ_V and η_V) produce spectra that are very close to scale invariant. Steeper potentials or potentials with significant curvature lead to a larger tilt.
As simple examples, monomial potentials $V(\phi) \propto \phi^p$ yield

$$n_s - 1 \simeq -\frac{p+2}{2N_\star}\,, \tag{12.79}$$

where $N_\star$ is the number of e-folds before the end of inflation when the pivot scale $k_\star$ exits the Hubble radius. Plateau–type potentials, where the potential becomes flat far away from the minimum, instead give $n_s - 1 \simeq -2/N_\star$, which for $N_\star \sim 50$–60 naturally leads to $n_s \simeq 0.96$–0.97 and therefore fits current data very well.

Running of the spectral tilt. In principle the spectral index itself can depend on scale. This scale dependence is quantified by the *running of the spectral tilt*, defined as the logarithmic derivative of $n_s(k)$ with respect to k,

$$\alpha_s \equiv \left.\frac{{\rm d}n_s}{{\rm d}\ln k}\right|_{k=k_\star} = \left.\frac{{\rm d}^2 \ln \Delta_{\mathcal{R}}^2(k)}{{\rm d}(\ln k)^2}\right|_{k=k_\star}\,. \tag{12.80}$$

For an exact power law spectrum of the form (12.70), the tilt is constant in k and therefore $\alpha_s = 0$. In single–field slow–roll inflation the running is a second–order slow–roll effect. Using the slow–roll hierarchy one finds, in terms of the potential slow–roll parameters,

$$\alpha_s \simeq 16\epsilon_V\eta_V - 24\epsilon_V^2 - 2\xi_V^2\,, \tag{12.81}$$

where

$$\xi_V^2 \equiv M_{\rm Pl}^4 \frac{V_{,\phi}V_{,\phi\phi\phi}}{V^2}\,. \tag{12.82}$$

Since $\epsilon_V, \eta_V, \xi_V^2$ are all small in the slow–roll regime, the running is generically expected to be of order 10^{-3} or smaller.

Measurements The primordial spectrum is not observed directly, but through the matter power spectrum and the cosmic microwave background (CMB)., From precise observations of the CMB temperature, polarisation, and lensing data, the Planck satellite measured

$$n_s = 0.9649 \pm 0.0042 \quad (68\%\ \text{CL})\,, \tag{12.83}$$

at a pivot scale $k_\star = 0.05\,\text{Mpc}^{-1}$ [?]. The tilt is therefore detected at high significance and is red, $n_s < 1$, in good agreement with the expectation from slow–roll models with slightly decreasing H and increasing ϵ. CMB experiments also constrain the running by fitting $\Delta^2_{\mathcal{R}}(k)$ with a Taylor expansion around the pivot scale,

$$\ln \Delta^2_{\mathcal{R}}(k) = \ln A_s + (n_s - 1)\ln\left(\frac{k}{k_\star}\right) + \tfrac{1}{2}\,\alpha_s \left[\ln\left(\frac{k}{k_\star}\right)\right]^2 + \cdots\,, \tag{12.84}$$

Current data are consistent with $\alpha_s = 0$ within errors, with a preferred value of order a few $\times 10^{-3}$ and no statistically significant detection of running. This is compatible with the expectation from simple slow–roll inflationary potentials, which predict a nearly power–law spectrum with a small, second–order running.

It is important to stress that the CMB only probes a finite range of comoving scales. The primary CMB anisotropies are well measured for multipoles $2 \lesssim \ell \lesssim 2500$. These multipoles roughly correspond to comoving wavenumbers

$$k \sim \frac{\ell}{D_*} \simeq 10^{-4}\,\text{Mpc}^{-1} \text{ to } 0.2\,\text{Mpc}^{-1}\,, \tag{12.85}$$

where $D_* \simeq 14\,\text{Gpc}$ is the comoving distance to last scattering. On larger scales there are only very few independent modes on the sky (cosmic variance), while on smaller scales Silk damping and instrumental resolution suppress the observable signal. The measured value of n_s therefore constrains the primordial spectrum only over roughly three orders of magnitude in k around the pivot scale. Any statement about the behaviour of $P_{\mathcal{R}}(k)$ on much smaller or larger scales necessarily relies on extrapolating the simple power law (12.70) beyond the range that is directly constrained by the CMB.

From the inflationary point of view, the observed value $n_s \simeq 0.965$ provides a powerful constraint on the shape of $V(\phi)$ through (12.78). Together with bounds on the tensor–to–scalar ratio, this allows one to select viable families of potentials in the (ϵ_V, η_V) plane, and it shows that the inflaton potential must be very flat over the field range corresponding to the observable ~ 50–60 e-folds of inflation.

A General Relativity in a nutshell

In this lesson, we give a lightning review of the results in General Relativity (GR) that we will need in this class and we set our notation. The reader familiar with GR and keen to start with cosmology can skip this lesson move directly to FLRW spacetimes. In the following, we discuss the equivalence principle, geodesic equation, conservation of energy-momentum tensor and conserved charge densities.

In GR the well-known metric of Minkowski spacetime $\eta_{\mu\nu}$ is promoted to a dynamical metric field $g_{\mu\nu}$. The dynamics of $g_{\mu\nu}$ is determined by distribution of matter. The metric is not determined univocally, but only up to a choice of coordinates, which change the metric in a specific way (somewhat similar to gauge theories). The laws of nature formulated in GR are not only valid in every intertial frame, but in any frame whatsoever. The phenomenon of gravitation is described by postulating that object move along geodesics (shortest paths) of the metric. Additional dynamics due to non-gravitational forces is generalized from the usual Minkoskian expression to account for the coordinate covariance of the theory.
One can derive all General Relativity (GR) from two principles:

- The principle of equivalence of mass and inertia, a.k.a. the *equivalence principle*: free falling observers do not feel the effects of gravitation. Formally, in an open set around any spacetime point we can choose the *locally inertial frame* (LIF), namely coordinates such that the *metric tensor* is approximately Minkowski

$$g_{\mu\nu} = \eta_{\mu\nu}\,, \quad \text{and} \quad \partial_\gamma g_{\mu\nu} \equiv g_{\mu\nu,\gamma} = 0\,. \tag{A.1}$$

- The principle of *general covariance*[104]: equations must be invariant in form under a change of coordinates.

Strategy: first, write down the equations governing a (sufficiently small) system in the absence of gravity; second, re-write them in a covariant way. The equation is now valid in the presence of gravity, i.e. in any coordinates.[105]

Clocks, rods and tensors Take two spacetime points separated by an infinitesimal timelike interval. We call this a clock because there is a reference frame in which this is some observer proper time. To go from special to general relativity we just start from the right expression in the LIF and make it generally covariant:

$$dx^\mu dx^\nu \eta_{\mu\nu} \doteq -dT^2 \quad \rightarrow \quad dx^\mu dx^\nu g_{\mu\nu} = -dT^2\,, \tag{A.2}$$

where we use the signature $(-,+++)$. Similarly for length contraction consider an infinitesimal spacelike interval, aka a rod,

$$dx^\mu dx^\nu \eta_{\mu\nu} \doteq dL^2 \quad \rightarrow \quad dx^\mu dx^\nu g_{\mu\nu} = dL^2\,. \tag{A.3}$$

Clocks and rod are dilated and contracted in the presence of a gravitation field (unlike in special relativity).

[104] CFU: *Be sure to understand the difference between covariance and invariance*
[105] CFU: *Is a free falling elevator a locally inertial frame? (yes) Are we in this room in a LIF? (no) Is the moon in a LIF? (yes as a point particle, no because of small tides) the earth or the sun? (same as the moon)*

Covariant objects A covariant or contravariant *scalar*, *vector* and *tensor* transform under a change of coordinates $x' = x'(x)$ as

$$\boxed{\phi'(x') = \phi(x)\,, \quad v'^{\mu'} = \frac{\partial x'^{\mu'}}{\partial x^{\mu}} v^{\mu}\,, \quad g'_{\mu'\nu'} = \frac{\partial x^{\mu}}{\partial x'^{\mu'}}\frac{\partial x^{\nu}}{\partial x'^{\nu'}} g_{\mu\nu}}\,. \tag{A.4}$$

A trick to get it right is to put the prime both on the tensor and on the indices. A tensor that is zero in one frame is zero in every frame.[106][107] Normal derivatives do not in general transform as tensors (unless they act on a scalar), and need to be supplemented by a "connection" to transform covariantly. The covariant derivative ∇_μ, indicated also by the label $;\mu$ appended to tensor it acts on, is defined as follows

$$\begin{aligned}
\nabla_\mu A &= A_{;\mu} = \frac{\partial A}{\partial x^\mu} = \partial_\mu A = A_{,\mu}\,, && \text{(A.5)}\\
\Box A \equiv \nabla^\mu \nabla_\mu A &= A_{;\mu}^{;\mu} = \frac{1}{\sqrt{-g}}\partial_\mu\left(g^{\mu\nu}\sqrt{-g}\partial_\nu A\right)\,, && \text{(A.6)}\\
A^\mu_{;\nu} &= \frac{\partial A^\mu}{\partial x^\nu} + \Gamma^\mu_{\sigma\nu}A^\sigma\,, && \text{(A.7)}\\
A_{\mu;\nu} &= \frac{\partial A_\mu}{\partial x^\nu} - \Gamma^\sigma_{\mu\nu}A_\sigma\,, && \text{(A.8)}\\
A^\mu_{\sigma;\nu} &= \frac{\partial A^\mu_\sigma}{\partial x^\nu} - \Gamma^m_{\sigma\nu}A^\mu_m + \Gamma^\mu_{m\nu}A^m_\sigma\,, && \text{(A.9)}\\
A_{\mu\sigma;\nu} &= \frac{\partial A_{\mu\sigma}}{\partial x^\nu} - \Gamma^\rho_{\mu\nu}A_{\rho\sigma} - \Gamma^\rho_{\sigma\nu}A_{\mu\rho}\,, && \text{(A.10)}
\end{aligned}$$

where the Christoffel symbol

$$\tag{A.11}$$

$$\Gamma^\mu_{\alpha\beta} \equiv \frac{1}{2}g^{\mu\gamma}\left(g_{\alpha\gamma,\beta} + g_{\beta\gamma,\alpha} - g_{\alpha\beta,\gamma}\right)\,, \tag{A.12}$$

vanishes in the LIF[108]. Note for this form of the geodesic equation to be valid, u must be linearly related to the proper time (for a generic time parameter u and additional term appears in this equation, see e.g. [9]). It is useful to remember that

$$\Gamma^\mu_{\nu\rho} = \Gamma^\mu_{\rho\nu}\,, \quad \Gamma_{\mu\nu\rho} + \Gamma_{\nu\mu\rho} = g_{\mu\nu,\rho}\,, \quad \Gamma^\mu_{\mu\lambda} = \frac{1}{\sqrt{g}}\partial_\lambda\sqrt{g}\,, \tag{A.13}$$

and that $\Gamma^\mu_{\nu\rho}$ does *not* not transform as a tensor. Under a coordinate transformation $x^\mu \to y^\mu(x)$ one finds

$$\Gamma^\alpha_{\beta\gamma} = \Gamma^\mu_{\nu\rho}J^\alpha_\mu J^\nu_\beta J^\rho_\gamma + J^\alpha_\mu \partial_\beta J^\mu_\gamma\,, \quad J^\alpha_\mu \equiv \frac{\partial y^\alpha}{\partial x^\mu}\,. \tag{A.14}$$

[106] CFU: *What is the inverse of $g_{\mu\nu}$? ($g^{\mu\nu}$)*

[107] CFU: *What is the metric (tensor)? (An infinitesimal distance, i.e. the norm of tangent vectors, not the distance between points on the manifold. By integrating the norm of the tangent vector to some curve (computed with the metric tensor), we can compute the length of the curve. Define by transforming as a two tensor and being $\eta_{\mu\nu}$ in the LIF.)*

[108] CFU: *Why is it called "geodesic" equation? Because the solution minimizes $\int \sqrt{\partial_u x^\mu \partial_u x_\mu}du$*

Geodesic equation Derivative of dx^μ with respect to some fixed time u, such as the proper time along a trajectory dx^μ is a vector because only dx^μ changes, while u does not since it is uniquely defined by the timeline of dx^μ. But the second derivative is not a vector. We need another non-vector to make a covariant expression. From Newton's law in the LIF we find

$$\frac{d^2x^\mu}{du^2} \doteq 0 \quad \rightarrow \quad \boxed{\frac{d^2x^\mu}{du^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{du}\frac{dx^\beta}{du} = 0}\,. \tag{A.15}$$

The solution are called geodesics and *maximize* proper time for any timelike path connection two timelike events.

Lie derivatives This discussion is based on App. B of [12] and Ch. 8 of [9]. Consider a vector field V^μ on some manifold M, which we parameterise locally with coordinates x. The vector generates *integral curves*, i.e. solutions of

$$\frac{\partial x^\mu}{\partial t} = V^\mu(x)\,. \tag{A.16}$$

These curves are tangent to V^μ at every point. We can think of an integral curve $x^\mu(t)$ as a one parameter family of (finite) changes of coordinates $x^\mu(t_0) \rightarrow x'^\mu = x^\mu(t)$. Instead of a passive coordinate transformation in which tensors on the manifold remain fixed, and points of the manifold change name according to the change of coordinates above, we can define an active transformations in which we drop the prime from the new coordinates x' and impose a change of all the tensors with fixed coordinates, i.e. a diffeomorphism

$$x \rightarrow x'\,, \quad T(x) \rightarrow T(x') \quad \text{(passive change of coords)}\,, \tag{A.17}$$

$$T(x) \rightarrow T'(x') \equiv T(x(x')) \quad \text{(active diffeomorphism)}\,. \tag{A.18}$$

Then, we can ask how a given tensor changes under infinitesimal diffeomorphism generated by an integral curve. We define the Lie derivative $\mathcal{L}$ of any covariant tensor $T^{\cdots}_{\cdots}$ (i.e. transforming as in Eq. (A.4)), in the V^μ direction is given by

$$\mathcal{L}_V T^{\cdots}_{\cdots}(x) \equiv \lim_{\epsilon\rightarrow 0}\frac{T^{\cdots}_{\cdots}(x) - T'^{\cdots}_{\cdots}(x)}{\epsilon}\,, \quad \text{with} \quad x'^\mu(x) = x^\mu + \epsilon V^\mu(x)\,. \tag{A.19}$$

As the name suggests, this derivative is a linear operator and obeys the Leibniz rule

$$\mathcal{L}_V(aT + bS) = a\mathcal{L}_V T + b\mathcal{L}_V S\,, \quad \mathcal{L}_V\,(T * S) = (\mathcal{L}_V T) * S + T * (\mathcal{L}_V S)\,, \tag{A.20}$$

where $*$ represents any index contraction. For scalar, vector and tensor field one finds

$$\mathcal{L}_V \phi = V^\mu \partial_\mu \phi\,, \tag{A.21}$$

$$\mathcal{L}_V W^\mu = V^\nu\partial_\nu W^\mu - W^\nu \partial_\nu V^\mu = V^\nu \nabla_\nu W^\mu - W^\nu \nabla_\nu V^\mu\,, \tag{A.22}$$

$$\mathcal{L}_V W_\mu = V^\nu\partial_\nu W_\mu + W_\nu \partial_\mu V^\nu = V^\nu \nabla_\nu W^\mu + W_\nu \nabla_\mu V^\nu\,, \tag{A.23}$$

$$\mathcal{L}_V T_{\mu\nu} = V^\rho \nabla_\rho T_{\mu\nu} + T_{\rho\nu}\nabla_\mu V^\rho + T_{\mu\rho}\nabla_\nu V^\rho\,. \tag{A.24}$$

where the intermediate expressions makes it explicit that the Lie derivatives are independent of the metric. Notice that the Lie Derivative is still a tensor of the same rank as suggested by Eq. (A.19). For vectors $\mathcal{L}_V W = -\mathcal{L}_W V$. In particular, since the metric is symmetric and covariantly constant, one finds

$$\mathcal{L}_V g_{\mu\nu} = \nabla_\mu \epsilon_\nu + \nabla_\nu \epsilon_\mu \,. \tag{A.25}$$

Since isometries are defined by $g'(x) = g(x)$, this give the Killing equation for the generators of isometries

$$\nabla_\mu \epsilon_\nu + \nabla_\nu \epsilon_\mu = 0 \,. \tag{A.26}$$

Note that any linear combination of Killing vectors is itself a Killing vector and so generates isometries.

Riemann and Ricci In Euclidean spacetime, initially parallel geodesics, i.e. straight lines, remain forever parallel and never intersect. Conversely, the convergence or divergence of geodesics is a manifestation of the curvature of spacetime. Similarly, covariant derivatives on a curved spacetime do not commute and parallel transport along a closed loop does not leave a vector unchanged. The Riemann tensor quantifies the deviation from flat spacetime expectation

$$[\nabla_\mu, \nabla_\nu] V_\rho = R_{\rho\sigma\mu\nu} V^\sigma \,, \tag{A.27}$$

where the covariant tensor $R^\rho{}_{\sigma\mu\nu}$ is given by[109]

$$R^\rho{}_{\sigma\mu\nu} = \partial_{[\mu} \Gamma^\rho_{\nu]\sigma} + \Gamma^\rho_{[\mu\lambda} \Gamma^\lambda_{\nu]\sigma} \,, \tag{A.28}$$

with anti-symmetrization defined in (0.4). The follow symmetry properties are useful

$$R_{\rho\sigma\mu\nu} = R_{\mu\nu\rho\sigma} \,, \qquad\qquad R_{\rho\sigma\mu\nu} = -R_{\sigma\rho\mu\nu} = R_{\sigma\rho\nu\mu} \,, \tag{A.29}$$

$$R_{\rho\sigma\mu\nu} + R_{\rho\nu\sigma\mu} + R_{\rho\mu\nu\sigma} = 0 \,, \tag{A.30}$$

where the latter is also known as first (algebraic) Bianchi identity. The second Bianchi identities are instead a differential relation among the components of the Riemann tensor

$$\nabla_\lambda R_{\alpha\beta\mu\nu} + \nabla_\nu R_{\alpha\beta\lambda\mu} + \nabla_\mu R_{\alpha\beta\nu\lambda} = 0 \tag{A.31}$$

which can be derived from the Jacobi identities for the commutator (A.27) of covariant derivatives (see Sec 7.8 of [9]).
Two well known contractions of Riemann are the Ricci tensor and Ricci scalar[110],

$$R_{\mu\nu} \equiv R_{\rho\mu\rho\nu} \,, \quad R \equiv g^{\mu\nu} R_{\mu\nu} = R^\mu_\mu \,. \tag{A.32}$$

Contracting all but one of the indices in the Bianchi equations (A.31) with the metric, one gets a contracted Bianchi identity for the Einstein tensor $G_{\mu\nu}$

$$\nabla^\mu G_{\mu\nu} \equiv \nabla^\mu \left(R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \right) = 0 \,. \tag{A.33}$$

[109] CFU: *How many spacetime derivates acting on the metric appear in* ${}^{\rho}{}_{\sigma\mu\nu}$*?*
[110] CFU: *How do* ${}^{\rho}{}_{\sigma\mu\nu}$*,* $R_{\mu\nu}$ *and* R *change under a constant rescaling of the metric* $g_{\mu\nu\to\lambda g_{\mu\nu}}$*?*

Newtonian limit and gravitational time dilation This limit is relevant for the formation of Large Scale Structures (LSS). Consider slowly moving particles, namely $\partial_u x^i \ll \partial_u x^0$ and choose time to be proper time so that $\partial_u x^0 = 1$. Then the geodesic equation is simply

$$\frac{d^2 x^i}{du^2} = -\Gamma^i_{00} \,. \tag{A.34}$$

For weak gravity, $g = \eta + h$, we can expand to linear order in h. Assuming slow time dependence wrt the spacial dependence $\partial_0 h \ll \partial_i h$ one finds Newton's law of gravitation:

$$\ddot{x}^i = -\partial_i \phi \,, \tag{A.35}$$

with $g_{00} = -1 - 2\phi$ and ϕ the gravitational potential. This implies that proper time runs slower in the gravitational field of a planet ($\phi < 0$):

$$dT = \sqrt{-dx^\mu dx^\nu g_{\mu\nu}} = \sqrt{(1+2\phi) dx^0 dx^0} \simeq (1+\phi) dx^0 \,, \tag{A.36}$$

where dT is the proper time interval and dx^0 is some global time coordinate that we use to compare observers with different values of ϕ.

Einstein Equations and Energy-momentum tensor In GR the metric is dynamical and it's evolution is dictated by the EE's: the matter energy momentum tensor tells spacetime how to bend. If the matter theory is described by an action S, then [111] the energy momentum is given by (the sign depends on conventions)

$$\boxed{T^{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta S}{\delta g_{\mu\nu}}} \,. \tag{A.37}$$

If we limit ourselves to only two spacetime derivatives, there is only one covariant expression that reduces to Poisson equation:

$$\partial_i \partial^i \phi \doteq 4\pi G \rho \quad \to \quad \boxed{R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -8\pi G T_{\mu\nu} = -M_{\mathrm{Pl}}^{-2} T_{\mu\nu}} \,. \tag{A.38}$$

As a consequence $T^{\mu\nu}_{;\nu} = 0$. The conservation of energy ($\mu = 0$) and momentum ($\mu = i$) currents is given in GR by[112][113]

$$T^{\mu\nu}_{,\nu} \doteq 0 \quad \to \quad T^{\mu\nu}_{;\nu} \equiv T^{\mu\nu}_{,\mu} + \Gamma^\mu_{\kappa\nu} T^{\kappa\nu} + \Gamma^\nu_{\kappa\nu} T^{\kappa\mu} = 0 \,. \tag{A.39}$$

EE's can be derived from the Einstein Hilbert action (plus the Gibbons-Hawking-York boundary term which we omit here)

$$\boxed{S = \int d^4 x \sqrt{-g} \frac{M_{\mathrm{Pl}}^2}{2} (R + \Lambda)} \,, \tag{A.40}$$

where Λ is a *cosmological constant.*

[111] CFU: *How do we know this formula is right? It's covariant and in the comoving LIF* $u^\mu = (1, 0, 0, 0)$
[112] CFU: *What is* $g_{\mu\nu;\gamma}$*? Why? It vanishes in the LIF, therefore it vanishes everywhere*
[113] Remember that $X_{;\mu} \equiv \nabla_\mu X$ for any tensor X.

B A toolkit to study an equation

In every subject there are a few pivotal equations that needs to be understood as well as possible. Here we collect a step-by-step toolkit to study a given equation for the first time, with the goal of understanding its many implications. A partial, semi-ordered list of things to do contains:

1. **Form** Stare at the equation as you would stare at a beautiful painting. Take at least 30 seconds to just look at it. Discover all of its tiny indices, hidden dependences, overall form. Is it an algebraic or differential equation? If differential, to what order? Is it partial or ordinary?

2. **Variables** Enumerate and characterize the variables in the equations: what are they functions of, how do they appear (e.g. with or without derivatives, integrated over, implicitly, ...)

3. **Dimensional analysis** Know/review the mass dimension (or other dimension is $\hbar \neq 1 \neq c$) of every single parameter, variable and function appearing in the equation. Be sure to master this.

4. **Symmetries** Discuss the symmetries of the equation: is it covariant (i.e. invariant in form) under change of coordinates? is it exactly/approximately invariant under some other symmetry? How do you build new solutions from known ones?

5. **Limits** Enumerate simple limits in which the equation takes a simple, well-known or intuitive form or in which you know a (simple) solution

As an example, let me discuss the geodesic equation,

$$\frac{d^2x^\mu}{du^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{du}\frac{dx^\beta}{du} = 0\,. \tag{B.1}$$

1. **Form** Four second order partial differential equations for four variables $x^\mu(u)$, with two terms and Lorentz indices. Γ is evaluated at x^μ, and therefore depends implicitly on it.

2. **Variables** The particle spacetime position $x^\mu(u)$ as function of *proper* time u (or an affine transformation thereof $u' = \lambda u + c$). Γ are the Christauwful symbols, related to the metric and its first derivative as in Eq. (A.12). x^μ appears explicitly only with (time) derivatives (one or two derivatives), but it may appear without derivative inside Γ, if e.g. the metric is not translation invariant. The metric appears both without derivatives and with one derivative. In typical applications, the metric determines the “background” and it is not a “dynamical” variable in this equation.

3. **Dimensional analysis** $[x^\mu, u] = M^{-1}$, $[\Gamma] = M^1$, $[g_{\mu\nu}, g^{\mu\nu}] = M^0$. Each term in the equation is an overall M^{-1}.

4. **Symmetries** The full equation is covariant under general spacetime diffs. Only dx^μ/du is covariant, while d^2x^μ/du^2 and Γ are not. The two terms are not separately covariant. u is proper time and therefore invariant under diffs. The theory is *not* invariant under a general reparameterization of the particle worldline $u' = u'(u)$.

5. **Limits** In the local inertial frame (which always exists thanks to the equivalence principle), the geodesic equation becomes simply $\ddot{a} = 0$. In the Newtonian limit (see Sec. **??**), one finds $\ddot{a}_i = -\partial_i \phi$, as it should be.

C Lesson references and further reading

Cosmology There are many good introductory textbooks to cosmology. We especially like those by Scott Dodelson [23], Viatcheslav Mukhanov [46] and Steven Weinberg [68]. Where possible we follow Weinberg's notation.

Sec. 1 The discussion of isometries and FLRW spacetime follows Ch. 13 of Weinberg's old book [64]. The rest is very standard.

Sec. 2 The discussion of distances follows 2.2 of Dodelson. Curvature is discussed following 1.3.1 of Mukhanov.

Sec. 3 Further details can be found in specialized reviews: for dark energy see [16, 59]; for neutrinos see [24, 33, 37, 38]; for dark matter see [8].

Sec. 4 The horizon and flatness problems can be found in any textbook. The discussion of coherent superHubble perturbations was inspired by [22], while that of scale invariance borrows from [17] and [5]. A nice introductory discussion of dS and conformal diagram is given in Sec 1.3.6 and Sec. 2.3 of [46]. A more advanced discussion including QFT and Quantum Gravity in dS can be found in [61].

Sec. 5 The general discussion of inflation and slow-roll parameters can be found in any textbooks.

Sec. 6 Thermal history is summarized in most textbook, see e.g. Mukhanov 3 and expecially 3.2. The discussion of the Boltzmann equation follows closely 3.1 of Dodelson.

Sec. 7 In the Part III course we cover only BBN, but we leave here some material on recombination and dark matter decoupling. They all follow closely 4 of Dodelson.

Sec. 8 In these lecture notes we have mostly followed Weinberg's book [68]. The equivalent chapter in Dodelson's book is Ch. 5. Two classic references on Cosmological Perturbation Theory are the review by Sasaki and Kodama [36] and that by Mukhanov, Feldman and Brandenberger [47].

App. A This Lesson follows App. B of Weinberg's book, Sections 2.1, 2.3 of Dodelson and selected topics from Blau's notes and Carrol's book.

Box A.1 Conserved currents and charges Symmetries of the law of physics are transformations that commute with the time evolution and generate new solutions from old ones. Mathematically, we represent symmetries by transformations of the (field) variables that leave the action invariant. By Noether theorem, for each such symmetry, there is an associated conserved current $\partial_\mu J^\mu \doteq 0$. The corresponding covariant expression is clearly $\nabla_\mu J^\mu = 0$. But the distinction is actually irrelevant (as long as one is careful with the convention she is using) since for every covariantly conserved current J^μ, with $J^\mu_{;\mu} = 0$, one can define a normally conserved current $\tilde{J}^\mu \equiv \sqrt{-g} J^\mu$, since

$$\tilde{J}^\mu_{,\mu} = \sqrt{-g} J^\mu_{;\mu} = 0 \,. \tag{A.41}$$

The conserved charge $\dot{Q} = 0$ is then defined as usual by

$$Q \equiv \int d^3x \tilde{J}^\mu n_\mu = \int d^3x \sqrt{g} J^\mu n_\mu \,, \tag{A.42}$$

where the integral is over some spatial hypersurface defined by the perpendicular vector n^μ. It is always possible and sometimes useful to split a current as $J^\mu = \rho u^\mu$ with a normalised velocity u^μ and a density ρ:

$$\rho \equiv \sqrt{-J_\mu J^\mu} \,, \quad u^\mu \equiv \frac{J^\mu}{\rho} \quad \Rightarrow \quad u^\mu u_\mu = -1 \,. \tag{A.43}$$

The density ρ is a density per proper volume, which is transformed into a density per coordinate volume by the $\sqrt{-g}$ factor in (A.41).
Lorentz symmetries are special. They lead to the covariant conservation of the energy momentum tensor $T_{\mu\nu}$, see (A.39), but this cannot in general be used to define a conserved charge because the trick in (A.41) does not work for a two-tensor,

$$0 = \sqrt{-g} T^\mu_{\nu;\mu} = (\sqrt{-g} T^\mu_\nu)_{,\mu} + \Gamma^\nu_{\mu\lambda} T^{\mu\lambda} \,, \tag{A.44}$$

and we are stuck with the last term. Energy and momentum are globally conserved iff there exist a Killing vector ϵ, which satisfies (A.26), $\epsilon_{(\mu;\nu)} = 0$. Then one can build the covariantly conserved current $J^\mu_\epsilon \equiv T^\mu_\nu \epsilon^\nu$, which has only one index, and proceed as above. In cosmology we will be interested in homogeneous and isotropic spacetimes, with six space-like Killing vectors but no time-like Killing vector. As a consequence we can define some globally conserved momentum and angular momentum, but no globally conserved energy. Intuitively, the cosmological spacetime exchanges energy with any system living on it, injecting and subtracting energy depending on the dynamics.

References

[1] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. 2015.

[2] P. A. R. Ade et al. Planck 2015 results. XVII. Constraints on primordial non-Gaussianity. Astron. Astrophys., 594:A17, 2016.

[3] James Aguirre et al. The Simons Observatory: Science goals and forecasts. 2018.

[4] Nima Arkani-Hamed, Hsin-Chia Cheng, Markus A. Luty, and Shinji Mukohyama. Ghost condensation and a consistent infrared modification of gravity. JHEP, 05:074, 2004.

[5] Nima Arkani-Hamed and Juan Maldacena. Cosmological Collider Physics. 2015.

[6] Erik Aver, Keith A. Olive, and Evan D. Skillman. The effects of He I ?10830 on helium abundance determinations. JCAP, 1507(07):011, 2015.

[7] James M Bardeen. Gauge-invariant cosmological perturbations. Physical Review D, 22(8):1882, 1980.

[8] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. Phys. Rept., 405:279–390, 2005.

[9] Matthias Blau. Lecture notes on General Relativity. page 951.

[10] Dillon Brout et al. The Pantheon+ Analysis: Cosmological Constraints. Astrophys. J., 938(2):110, 2022.

[11] T. S. Bunch and P. C. W. Davies. Quantum Field Theory in de Sitter Space: Renormalization by Point Splitting. Proc. Roy. Soc. Lond. A, 360:117–134, 1978.

[12] Sean M. Carroll. Spacetime and Geometry: An Introduction to General Relativity. Cambridge University Press, 7 2019.

[13] Clifford Cheung, Paolo Creminelli, A. Liam Fitzpatrick, Jared Kaplan, and Leonardo Senatore. The Effective Field Theory of Inflation. JHEP, 03:014, 2008.

[14] Sidney R. Coleman and Frank De Luccia. Gravitational Effects on and of Vacuum Decay. Phys. Rev. D, 21:3305, 1980.

[15] Ryan Cooke, Max Pettini, Regina A. Jorgenson, Michael T. Murphy, and Charles C. Steidel. Precision measures of the primordial abundance of deuterium. Astrophys. J., 781(1):31, 2014.

[16] Edmund J. Copeland, M. Sami, and Shinji Tsujikawa. Dynamics of dark energy. Int. J. Mod. Phys., D15:1753–1936, 2006.

[17] Paolo Creminelli. Conformal invariance of scalar perturbations in inflation. Phys. Rev. D, 85:041302, 2012.

[18] Paolo Creminelli. Conformal invariance of scalar perturbations in inflation. Phys. Rev., D85:041302, 2012.

[19] Paolo Creminelli, Markus A. Luty, Alberto Nicolis, and Leonardo Senatore. Starting the Universe: Stable Violation of the Null Energy Condition and Non-standard Cosmologies. JHEP, 12:080, 2006.

[20] Paolo Creminelli, Jorge Norena, and Marko Simonovic. Conformal consistency relations for single-field inflation. JCAP, 1207:052, 2012.

[21] Bryce S. DeWitt. Quantum Theory of Gravity. 1. The Canonical Theory. Phys. Rev., 160:1113–1148, 1967.

[22] Scott Dodelson. Coherent phase argument for inflation. AIP Conf. Proc., 689:184–196, 2003. [,184(2003)].

[23] Scott Dodelson. Modern Cosmology. Academic Press, Amsterdam, 2003.

[24] A. D. Dolgov. Neutrinos in cosmology. Phys. Rept., 370:333–535, 2002.

[25] S. Dubovsky, T. Gregoire, A. Nicolis, and R. Rattazzi. Null energy condition and superluminal propagation. JHEP, 03:025, 2006.

[26] Bernardo Finelli, Garrett Goon, Enrico Pajer, and Luca Santoni. Soft Theorems For Shift-Symmetric Cosmologies. Phys. Rev., D97(6):063531, 2018.

[27] D. J. Fixsen. The Temperature of the Cosmic Microwave Background. Astrophys. J., 707:916–920, 2009.

[28] D. J. Fixsen, E. S. Cheng, J. M. Gales, John C. Mather, R. A. Shafer, and E. L. Wright. The Cosmic Microwave Background spectrum from the full COBE FIRAS data set. Astrophys. J., 473:576, 1996.

[29] T. Garidi. What is mass in de Sitterian physics? 9 2003.

[30] Jaume Garriga and Viatcheslav F. Mukhanov. Perturbations in k-inflation. Phys. Lett., B458:219–225, 1999.

[31] M. Gervasi, M. Zannoni, A. Tartari, G. Boella, and G. Sironi. TRIS II: search for CMB spectral distortions at 0.60, 0.82 and 2.5 GHz. Astrophys. J., 688:24, 2008.

[32] Alan H. Guth. The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. Phys. Rev., D23:347–356, 1981.

[33] Steen Hannestad. Primordial neutrinos. Ann. Rev. Nucl. Part. Sci., 56:137–161, 2006.

[34] J. B. Hartle and S. W. Hawking. Wave Function of the Universe. Phys. Rev. D, 28:2960–2975, 1983.

[35] Kurt Hinterbichler, Lam Hui, and Justin Khoury. Conformal Symmetries of Adiabatic Modes in Cosmology. JCAP, 1208:017, 2012.

[36] Hideo Kodama and Misao Sasaki. Cosmological Perturbation Theory. Prog. Theor. Phys. Suppl., 78:1–166, 1984.

[37] Julien Lesgourgues and Sergio Pastor. Massive neutrinos and cosmology. Phys. Rept., 429:307–379, 2006.

[38] Julien Lesgourgues and Sergio Pastor. Neutrino mass from Cosmology. Adv. High Energy Phys., 2012:608515, 2012.

[39] Andrew R. Liddle, Paul Parsons, and John D. Barrow. Formalizing the slow roll approximation in inflation. Phys. Rev., D50:7222–7232, 1994.

[40] David H. Lyth and Antonio Riotto. Particle physics models of inflation and the cosmological density perturbation. Phys. Rept., 314:1–146, 1999.

[41] Chung-Pei Ma and Edmund Bertschinger. Cosmological perturbation theory in the synchronous and conformal Newtonian gauges. Astrophys. J., 455:7–25, 1995.

[42] M. Maltoni, T. Schwetz, M. A. Tortola, and J. W. F. Valle. Status of global fits to neutrino oscillations. New J. Phys., 6:122, 2004.

[43] G. Mangano, G. Miele, S. Pastor, and M. Peloso. A Precision calculation of the effective number of cosmological neutrinos. Phys. Lett., B534:8–16, 2002.

[44] Jerome Martin, Christophe Ringeval, and Vincent Vennin. Encyclopaedia Inflationaris. Phys. Dark Univ., 5-6:75–235, 2014.

[45] Charles W. Misner, K. S. Thorne, and J. A. Wheeler. Gravitation. W. H. Freeman, San Francisco, 1973.

[46] V. Mukhanov. Physical Foundations of Cosmology. Cambridge University Press, Oxford, 2005.

[47] Viatcheslav F. Mukhanov, H. A. Feldman, and Robert H. Brandenberger. Theory of cosmological perturbations. Part 1. Classical perturbations. Part 2. Quantum theory of perturbations. Part 3. Extensions. Phys. Rept., 215:203–333, 1992.

[48] P. James Peebles. The Large-Scale Structure of the Universe. Princeton University Press, 11 1980.

[49] S. Perlmutter et al. Measurements of Omega and Lambda from 42 high redshift supernovae. Astrophys. J., 517:565–586, 1999.

[50] Cyril Pitrou, Xavier Roy, and Obinna Umeh. xPand: An algorithm for perturbing homogeneous cosmologies. Class. Quant. Grav., 30:165002, 2013.

[51] Joseph Polchinski. Effective field theory and the Fermi surface. In Theoretical Advanced Study Institute (TASI 92): From Black Holes and Strings to Particles Boulder, C pages 0235–276, 1992.

[52] Adam G. Riess et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. Astron. J., 116:1009–1038, 1998.

[53] Adam G. Riess et al. A 2.4% Determination of the Local Value of the Hubble Constant. Astrophys. J., 826(1):56, 2016.

[54] V. A. Rubakov. Phantom without UV pathology. Theor. Math. Phys., 149:1651–1664, 2006. [Teor. Mat. Fiz.149,409(2006)].

[55] A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. Pisma Zh. Eksp. Teor. Fiz., 5:32–35, 1967.

[56] Santiago Agüí Salcedo, Thomas Colas, Lennard Dufner, and Enrico Pajer. An Open System Approach to Gravity. 7 2025.

[57] D. S. Salopek and J. R. Bond. Nonlinear evolution of long wavelength metric fluctuations in inflationary models. Phys. Rev. D, 42:3936–3962, 1990.

[58] Misao Sasaki and Ewan D. Stewart. A General analytic formula for the spectral index of the density perturbations produced during inflation. Prog. Theor. Phys., 95:71–78, 1996.

[59] Alessandra Silvestri and Mark Trodden. Approaches to Understanding Cosmic Acceleration. Rept. Prog. Phys., 72:096901, 2009.

[60] D. T. Son. Low-energy quantum effective action for relativistic superfluids. 2002.

[61] Marcus Spradlin, Andrew Strominger, and Anastasia Volovich. Les Houches lectures on de Sitter space. In Proceedings, NATO Advanced Study Institute, Euro Summer School, 76th session, Les Houches, France, pages 423–453, 2001.

[62] Kandaswamy Subramanian. The origin, evolution and signatures of primordial magnetic fields. Rept. Prog. Phys., 79(7):076901, 2016.

[63] David Wands. Local non-Gaussianity from inflation. Class. Quant. Grav., 27:124002, 2010.

[64] Steven Weinberg. Gravitation and Cosmology. John Wiley and Sons, New York, 1972.

[65] Steven Weinberg. Adiabatic modes in cosmology. Phys. Rev., D67:123504, 2003.

[66] Steven Weinberg. Damping of tensor modes in cosmology. Phys. Rev., D69:023503, 2004.

[67] Steven Weinberg. The Quantum theory of fields. Vol. 1: Foundations. Cambridge University Press, 2005.

[68] Steven Weinberg. Cosmology. 2008.