Notes on Theoretical Cosmology and Structure Formation

Emanuele Francesco Besana

Abstract

The following notes contain a comprehensive, but hopefully self-contained, exposition on the topics of theoretical cosmology and structure formation. The aim is to clarify many concepts that are often treated vaguely in the literature, providing a coherent and understandable explanation. These notes are an effort by the author to summarize his knowledge on these topics, and were partly developed during the writing of his master's thesis. The concepts are primarily introduced and explained from a theoretical perspective, reflecting the author's inclination. Consequently, there are few to no mentions of experimental aspects.

References

The author has made use of the following material in the writing of these notes:

Textbooks

- D. Baumann, *Cosmology*, Cambridge University Press, 7 2022
- S. Dodelson and F. Schmidt, *Modern Cosmology*, 2020
- S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, New York: John Wiley and Sons, 1972
- S. Weinberg, *Cosmology*, 2008
- M. Gasperini, *Lezioni di Cosmologia Teorica*, UNITEXT, Milano: Springer Milan, 2012
- M. Maggiore, *Gravitational Waves, Vol. 2: Astrophysics and Cosmology*, Oxford University Press, 3 2018
- H. Mo and F.C. Van den Bosch and S. White, *Galaxy Formation and Evolution*, 2010

Notes

- A. Riotto, *Lecture Notes on Cosmology*, 2013
- K. Suonio, *Cosmological Perturbation Theory I*, 2024

Papers

- M. Cirelli and A. Strumia and J. Zupan, *Dark Matter*, 6 2024
- S. Chabanier and M. Millea and N. Palanque-Delabrouille, *Matter power spectrum: from Lyα forest to CMB scales*, Mon. Not. Roy. Astron. Soc., vol. 489, no. 2, pp. 2246-2253, 2019
- A. R. Zentner, *The Excursion Set Theory of Halo Mass Functions, Halo Clustering, and Halo Growth*, Int. J. Mod. Phys. D, vol. 16, pp. 763-816, 2007
- T. M. C. Abbott et al., *The Dark Energy Survey: Cosmological Results With $\sim$ 1500 New High-Redshift Type Ia Supernovae Using the Full 5-year Dataset*, 1 2024
- T. Padmanabhan, *Aspects of gravitational clustering*, 11 1999

Contents

Part 1: The Homogeneous Universe

Chapter 1

The Standard Model of Cosmology

The study of the Universe, and the subject of Cosmology with it, began with some incredible observations. From then on, the constant interplay between theory and observations led the physics community to devise the "Standard Model of Cosmology" (or the "ΛCDM model", or the "Big Bang model"), which is to date the most successful paradigm when it comes to a description of the Universe.
The model is based on a synergy of Einstein's theory of General Relativity, which sets the equations for the dynamical evolution of the various components that make up the Universe, and the theory of inflation, usually in the form of some Quantum Field theory, that sets the conditions on the initial time hypersurface.
The ΛCDM model has made predictions that have seen remarkable validation from observations. Some of these successes are: the abundances of light elements predicted via Big Bang Nucleosynthesis are in striking accord with observations, the measurements of temperature and polarization of the Cosmic Microwave Background agree embarrassingly well with the theory, as well as correct predictions of multiple probes regarding large-scale structures in the Universe.

1.1 Large Scale Structure

In this section, we will touch on the early observations regarding the larger scale structure of the Universe.

Homogeneity and Isotropy

There is very good reason to believe that our Universe is spatially homogeneous and isotropic at very large scales. These are arguably the two most important attributes of the Universe, because they ensure that observations made from any vantage point (like the point in space where Earth is located and experiments are carried out), are representative of the whole Universe. This is clearly very powerful when it comes to testing our theories, made from our standpoint, against the properties of the distant Universe. These two conditions, homogeneity and isotropy, are together known as the Cosmological Principle, which remained an intelligent guess, made by theorists in order to simplify the mathematics, for most of the twentieth century. Evidence finally came forth at the end of the century when, for the first time, incredible evidence for the isotropy of the Universe was found via measurements of the temperature field of the Cosmic Microwave Background (CMB):

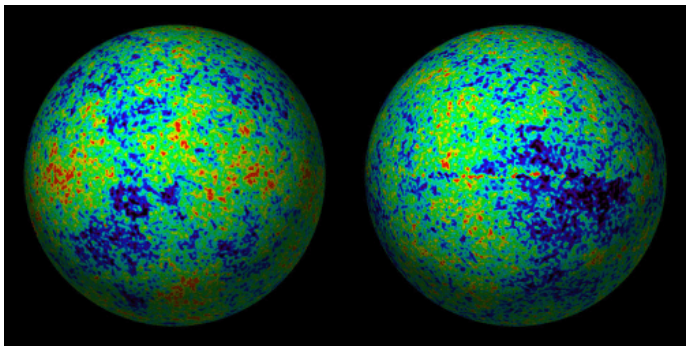

FIGURE 1.1: The cosmic microwave background as seen in the sky. The colormap distinguishes hot spots (red) from cold spots (blue). Note that the difference in temperature is of the order of 10^{-5} K.

Notwithstanding for the moment the origin of this background radiation, we only note that it is the most perfect black body found in nature, at a temperature of $\bar{T}_0 = 2.73$ K, and that it was created around 200,000 years after the Big Bang. More importantly, this temperature is exactly the same (except minor fluctuations of the order of 10^{-5} K) in every direction one looks.
The remarkable uniformity of the CMB indicates that the Universe, 200,000 years after the Big Bang, was isotropic and homogeneous to a very high degree of precision when observed at sufficiently large scales, up to the furthest observable distance that defines the observable Universe, which is around 3000 Mpc. In fact, in addition to the CMB, large redshift surveys, like the 2dF (2 degree field) survey, computed the position and distance of a huge number of galaxies, and found them to be distributed homogeneously and isotropically at scales $\gtrsim 100$ Mpc:

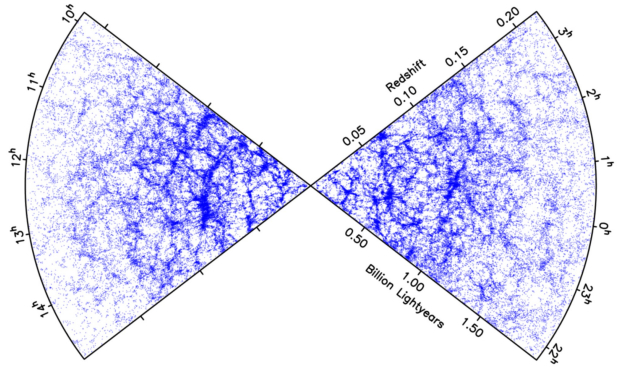

FIGURE 1.2: The 2dF galaxy survey

While on smaller scales there exist large inhomogeneities, such as galaxies, clusters and superclusters, limiting the validity of the Cosmological Principle to a limited range of large scales, the theory of inflation, as will be explained later, predicts that the Universe continues to be homogeneous and isotropic well beyond the present observable patch.

Spatial Curvature

Another important pillar of observational Cosmology regards the spatial geometry of the Universe. From measurements of the properties of the CMB (indeed, this background radiation is a rich source of information, as we will be able to appreciate even later on), one is able to distinguish between different type of spatial geometries. The reason for this is, roughly, because the photons of the CMB, moving along geodesics specified by the given spatial curvature, leave distinct imprints on the properties of the CMB. Current measurements reveal that the Universe appears to be flat, with no hint of any spatial curvature.

Hubble Diagram

Hubble, in 1929, first observed that galaxies moved away from our own at speeds proportional to their distance, i.e. the farther they were, the faster they moved. Technically, he noticed a redshift in emission spectra coming from the galaxies. This is usually regarded as the first evidence of the fact that our Universe is expanding. Hubble presented his findings in a "Hubble diagram":

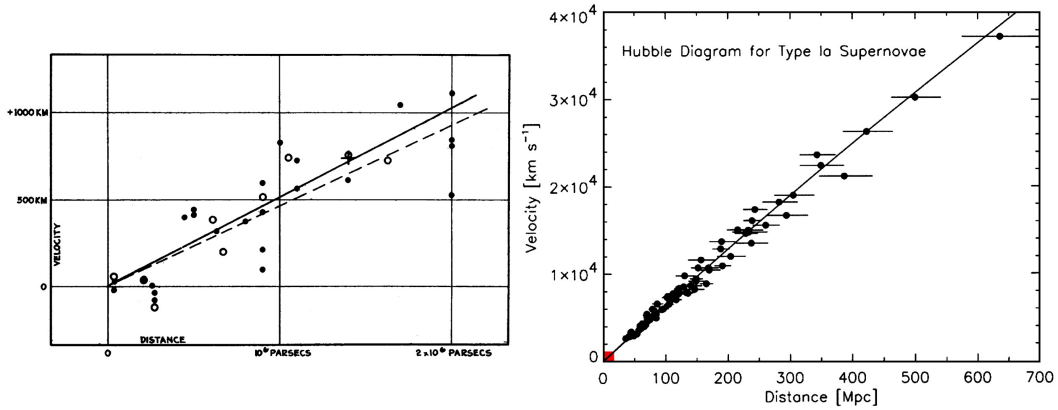

FIGURE 1.3: *Left*: Hubble's original results from his article in 1929. *Right*: the same diagram, but with more data points

The velocity, therefore, increases linearly with distance, with a slope given by H_0:

$$v_{\text{gal}} = H_0 D \tag{1.1}$$

where D is the distance of the galaxy, and H_0 is dubbed the Hubble constant, that we will encounter and explain in more details later on. Equation (1.1) is universally known as Hubble's law for the recession of galaxies, and should be read as a distance-redshift relation, since the velocity v that appears in the equation is deduced from a measure of redshift. Despite the fact that this observation was made from Earth, large scale homogeneity implies that it should be valid for every observer, so that the Universe's spatial slices are expanding at every point on their surfaces.
Hubble's law, as will be much clearer below, is only a low-distance approximation, and the more rigorous result for the distance-redshift relation will be in the form of a power series, where the terms depend on the energy content of the Universe. Furthermore, the notion of redshift in an expanding Universe will need clarification.

1.2 The Energy Content of the Universe

Observations suggest that the mass-energy content is divided only into three major components, which we now care to analyze in detail.

Matter and the Big Bang Nucleosynthesis

While we will precisely determine what we mean by "matter" only later on, we can say here that cosmologists distinguish between:

- Baryonic matter: standard cosmological conventions dictate that all ordinary matter fall in this category, even though, for example, electrons are leptons. However, their mass is negligible compared to protons, neutrons and nuclei, so that the true baryons yield the biggest contribution to the mass of ordinary matter. Baryons make up stars, planets and every living organism, and account for only 5% of the energy content of the Universe. Furthermore, there is no substantial amount of anti-matter;

- Dark matter: a type of matter of which the nature still remains a mystery. There are many candidates for dark matter (such as right handed neutrinos, axions, primordial black holes etc...), but it is not clear as to whether it is comprised of one or multiple components. This type of matter constitutes 27% of the matter content of the Universe.

Evidence for dark matter on macroscopic scales was first found by Zwicky around 1933 while studying the properties of the galaxies in the Coma cluster, a galaxy cluster. A large discrepancy was found between the baryonic mass of the system estimated via the observed luminosity, and that calculated via considerations based on the virial theorem, which takes into account all possible gravitating mass. The suspicion of a non-luminous type of matter was therefore the most trivial explanation.
A few decades later, in the 1960s and 70s, Vera Rubin and collaborators found further evidence for dark matter while studying the rotation curves of galaxies: measurements of rotation speeds of hydrogen gas in the outer legs of galaxies, determined by inspecting the 21 cm line of hydrogen and converting from a blue/red-shift of the spectrum to the velocity of the gas. Rotation curves were found to saturate in the peripheries of galaxies, in stark disagreement with predictions based on Newtonian gravity if baryonic mass was assumed to be the main contributor to the mass of the galaxies.
Finally, macroscopic evidence for a non-interacting and non-baryonic type of matter was extracted from gravitational lensing observations of the bullet cluster, formed by two clusters that are thought to have previously collided. The major components of the cluster, stars and dark matter, behave rather differently during collision, allowing them to be studied separately. The stars, i.e. the baryonic matter, were only gravitationally slowed but not otherwise altered during the collision. In contrast, dark matter was detected indirectly by the gravitational lensing of background objects. The lensing was found to be strongest in two separated regions near the visible galaxies, providing support for the idea that most of the gravitation in the cluster resides in two regions of dark matter, which bypassed the gas regions during the collision. This new type of matter was hence called "dark" matter, reflecting its seemingly negligible interaction with baryonic matter or radiation.

Today, the strongest evidence of dark matter is microscopic in nature, and comes from the gravitational lensing of the Cosmic Microwave Background (CMB). The abundances of light elements, like hydrogen and helium, that make up the baryonic matter in the Universe, can be estimated via an accurate theory called Big Bang Nucleosynthesis (BBN). The total composition of baryonic matter is about 75% hydrogen, 25% helium, plus small amounts of heavier elements. The baryon energy density predicted theoretically via BBN clashed with observations of the CMB. As the photons travel through the Universe, the intervening large scale structure curves their paths, giving rise to the cold and hot spots in the CMB. Because the degree of lensing depends on the total amount of matter, we can compare this density with the abundance of baryons created via BBN. The results clearly indicated an asymmetry between the two types of matter.
There is therefore, at the present level of knowledge, agreement between very different probes of the Universe that the total matter density is comprised of only 30% of baryons, with 80% being in the form of dark matter.

Radiation

Cosmological conventions label as radiation every particle whose mass is negligible compared to its momentum, an example clearly being photons. The majority of the radiation contribution to the total energy density of the Universe is in the form of the CMB, whereas light emitted from extended objects, such as stars, only plays a minor role.
An important quantity in Cosmology measures the ratio of the number density of baryons to that of photons:

$$\eta = \frac{n_b}{n_\gamma} \simeq 10^{-9} \tag{1.2}$$

and is called the "baryon to photon ratio". It indicates that in the Universe there is roughly one baryon per 10^9 photons.
Finally, in general, very light particles are considered radiation as well, like neutrinos, whose mass actually results in a non-negligible effect on the evolution of galaxies. In total, the radiation makes up roughly 0.01% of our Universe, a rather small percentage.

Dark Energy and the Acceleration of the Universe

A multitude of independent evidence has now accumulated to indicate the existence of dark energy, a substance with an exotic equation of state which doesn't dilute with the expansion of the Universe.
One of the strongest pieces of evidence was found at the end of the 1990s, when observations of distances of Type Ia supernovae (SNe) yielded proof for an accelerating Universe. In other words, the Hubble diagram of in Figure (1.3) was extended to higher redshifts using a distance-redshift relation based on the luminosity of the sources, in this case supernovae. As was remarked above, the luminosity distance, as a function of redshift, generalizes Hubble's law and depends on the contents of the Universe, in particular on the total energy density. These observations then showed that SNe appeared fainter than predicted with a matter-only Universe, and the best fit required the inclusion of dark energy:

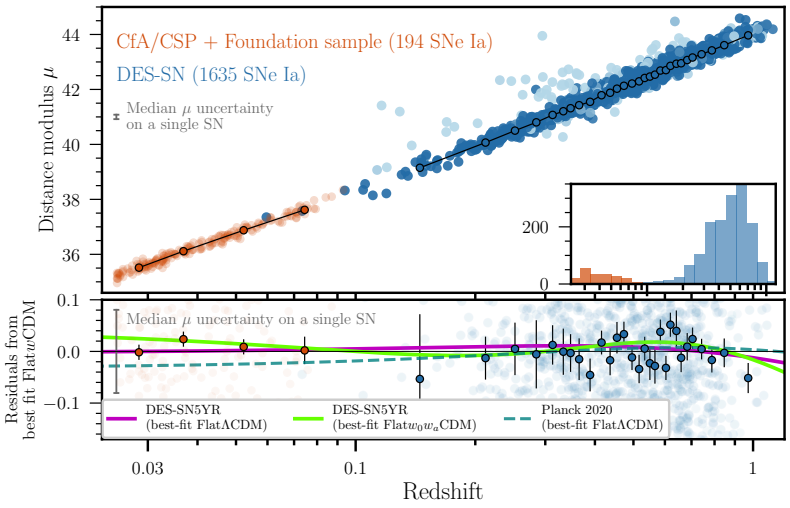

FIGURE 1.4: Hubble diagram for Type Ia supernovae taken by the Dark Energy Survey with 5 years of data (DES-SN5YR). The bottom panel shows the difference between the data and three models. The magenta line, which is the best fit ΛCDM model based on these results, unequivocally indicates the presence of dark energy.

On the vertical axis the distance modulus is shown, which is the logarithm of the luminosity distance. Best fit results of this curve, cross examined with CMB data, point to the concordance value of 70% of the energy density residing in this new dark energy.

1.3 The History of the Universe

Given that, today, the Universe is expanding, intuitively one can imagine that at earlier times objects must have been close to one another and the far past conditions very hot and dense. While this picture cannot be justified solely on the basis of the expansion of the Universe at present time, it is strongly supported by BBN measurements of the abundances of light elements. These measurements align closely with theoretical predictions, confirming the high-density and high-temperature conditions of the early Universe.
A standard practice in Cosmology is to describe the effect of the expansion by introducing a "scale factor" a, which measures the spatial extension of the Universe relative to its value at some specific time in its evolution. In this sense only ratios are physical and, as conventions dictate, a is set to 1 at present time. One of the effect of expansion is the stretching, proportional to the scale factor, of wavelengths emitted from distant objects. The result is a redshifting of photons, which is measured with a parameter called z, the "redshift", defined by:

$$1 + z = \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \frac{a_{\text{observed}}}{a_{\text{emitted}}} = \frac{1}{a_{\text{emitted}}} \tag{1.3}$$

The redshift is a neat way of organizing the locations (and times) at which objects reside, since the connection between a and z in (1.3) emphasizes that $z = 1$ is a time when the Universe was half the size it is now, $z = 2$ a third, $z = 3$ a fourth and so on. In Cosmology, then, different eras and events are denoted by their respective redshifts z. Correspondingly, $z = 0$ indicates the present time.

To understand the evolution of the Universe through its successive stages, one must determine the evolution of the scale factor, which, according to Einstein's equations, is governed by the evolution of the energy density of the Universe. Each component of the Universe, as introduced above, has a different equation of state, meaning their energy densities scale differently with the scale factor. Consequently, these components dominate the expansion of the Universe at different times. The results, computed through more detailed derivations, indicate that the Universe started off (after inflation, see later) in a period of radiation domination until $z \simeq 3400$, to which followed a long time of matter domination. Only later on, at $z \simeq 0.3$, did dark energy came to guide the evolution.
A useful equation relates the temperature of the Universe T to its age:

$$\frac{T}{1\ \text{MeV}} \approx \left(\frac{t}{1\ s}\right)^{-1/2} \tag{1.4}$$

meaning that, for instance, at $t = 1$s, the Universe had a temperature of about 1 MeV; moreover, the temperature scales with an inverse power of time, as expected for a Universe that expands in the future.
During primordial times, particles collided frequently and they were all in thermal equilibrium at a temperature T. Although the period of BBN is the farthest back that observations have probed the Universe and its equation of state, confirming this picture, it is only natural to take as initial conditions the Universe at around 100 GeV, where all the particles of the Standard Model were in equilibrium and therefore their abundances were roughly equal to one another. The rate of reactions, during early times, then, was extremely high, and a lot of interesting phenomena took place in a short timeframe. Shortly after the 100 GeV mark, in just 10^{-9} seconds the Universe expanded by a factor of 10^4 and the temperature dropped rapidly. During this short time the Universe went through successive evolutionary stages, the most important of which are listed here below.
At around 100 GeV, i.e. at $t = 10^{-11}$s, the electroweak (EW) symmetry of the Standard Model was broken in what is called the EW phase transition, where the weak and electromagnetic forces decoupled and particles acquired their mass. The detailed dynamics of this transition are still an open subject of research, even though the basics of it were confirmed by the discovery of the Higgs boson.
As the temperature drops the energy drops, so after some time the particle-antiparticle annihilation reaction is favored over the reverse process of particle creation. Clearly, since it requires less energy to form light particles rather than heavier ones, the first particles to disappear in this way were the most massive quarks, followed shortly after by the massive bosons W and Z, the tau lepton and the Higgs. Finally, at around 150 MeV, at $t = 10^{-5}s$, the remaining quarks, at this point having a very low kinetic energy, could condense into hadrons (mesons and protons/neutrons), resulting in the QCD phase transition.
Particles can fall out of thermal equilibrium when their interaction rate drops below the expansion rate of the Universe, in other words the Universe expands way too rapidly for the particles to interact. At that "moment" those particles will stop interacting with the environment, that is they decouple, creating a relic abundance. One of the most important decoupling events is the neutrino decoupling at around one second after the Big Bang, i.e. 1 MeV, which produced the Cosmic Neutrino Background (CνB).

Around 1 minute after the Big Bang, at $z \sim 10^8$, the temperature dropped enough to allow the creation of the first light nuclei Helium-4 and Lithium-7. This Big Bang Nucleosynthesis produced very few elements heavier than lithium, because there are no stable nuclei with 5 or 8 nucleons that would be required in order to sustain the reaction. Heavier elements instead formed only later on in the insides of stars, via stellar nucleosynthesis, where densities allow for the creation of, for instance, carbon, out of alpha particles via the 3α channel. Later, with the explosion of stars, these heavier elements, up until iron, spread throughout the Universe; instead, the origin of elements that lie beyond iron is still debated, but they most likely formed via explosive nucleosynthesis, either through s-processes during the last phases of stellar evolution, or through r-processes in supernovae explosions or neutron star mergers.
About 370000 years after the Big Bang, at $z = 1100$, the temperature dropped enough so that the first atoms could form. This process, known as recombination, when the photons stopped interacting with electrons (that in turn could begin to be captured by the nuclei), produced what is now detected as a relic photon background, the Cosmic Microwave Background, which was introduced above. Crucially, the CMB contains temperature fluctuations that, although very small ($\delta T/T \sim 10^{-5}$), contain very important information about the primordial Universe.

Inflation and Initial Conditions

If the cosmological principle held perfectly, if the distribution of matter in the Universe were homogeneous and isotropic at every scale, there would be no structures (galaxies, clusters etc...) at present time. Therefore, there needs to exist a mechanism which gives rise to these deviations from perfect uniformity.
Around and before the Planck time, $t \simeq 10^{-43}s$, it is expected that GR breaks down and gravity should become a quantum interaction. The nature of this quantum gravity era still remains speculative in the continued absence of a satisfactory quantum gravity theory. In fact, a model based only on an extrapolation of GR to higher energies has a number of conceptual problems when applied to the early Universe (see later). Inflation provides a framework for understanding the solutions to these problems, and it also predicts the generation of fluctuations, which are now believed to be responsible for the formation of all the large scale structure that we see. The same fluctuations are also responsible for the temperature anisotropies of the CMB. During inflation, the Universe is supposed to have gone through a rapid expansion over very little amount of time driven by one (or more) quantum field, the "inflaton".
The understanding of the Universe is still very far from complete and, despite the inflationary paradigm, cosmologists are still unable to predict the initial conditions for structure formation from first principles, but are forced to rely on a set of parameters taken from observational data.

The Growth of Structures

In any case, once the initial conditions have been specified, it is very straightforward to compute the way in which density perturbations evolve in time in a homogeneous

background. This is the realm of cosmological perturbation theory. When the universe is matter dominated, the perturbations will grow, because a region where the initial density is slightly higher than the mean will attract its sorroundings more strongly than average. Therefore overdense regions become more overdense over time, and underdense regions become more rarefied as matter will flow away from them. The exact rate at which perturbations grow will depend on the cosmological model, and in particular this growth will stop when dark energy comes to dominate the Universe. In any case, once a certain region is overdense enough, it will stop growing and start to collapse, which is the birth of large objects such as dark matter halos.
The first stars formed when the Universe was around 100 million years old, that is at $z = 25 - 6$. These were especially massive stars and, as such, lived for only a short amount of time, during which they emitted UV light that heated and ionized the surrounding gas. The dynamics of this process are called "reionization".
Over time, baryons start to fall in the potential wells created by dark matter halos, and form galaxies, which started to appear at $t = 10^9$ years, that is $z = 1 - 7/8$.

1.4 A Summary of the ΛCDM model

Any cosmological model worthy of being considered as the standard reference needs to be predictive and consistent with the facts that were established above. Summarizing, with regards to the large scale structure of the Universe, observations strongly suggest that:

- The Universe is spatially homogeneous and isotropic on scales that are $\gtrsim 100$ Mpc. Moreover, the spatial curvature of spatial slices is null, i.e. the Universe is flat;
- The Universe has been expanding since primordial times (at least since the epoch of BBN), and today it is doing so in an accelerated fashion;

Concerning the matter-energy content, on the other hand:

- The Universe is made up of about 30% of matter, baryonic and dark, negligible amounts of radiation and 70% of dark energy;
- A Cosmic Microwave Background at a temperature $\bar{T}_0 \simeq 2.73$ K pervades space. These photons comprise the totality of the radiation in the Universe and outnumber the baryons by 10^9 to 1;
- Fluctuations from homogeneity are present in both the matter fields, in the form of structures such as galaxies and clusters, and in the CMB, which displays temperature fluctuations of the order of 10^{-5} K.

The (flat) ΛCDM model is able to accommodate for all these facts as was delineated above. It introduces dark matter (CDM) as a natural additional density in the equations, where the nomenclature "CDM" stands for cold dark matter (as opposed to hot) and indicates that its average speed is negligible compared to the speed of light. Cold is preferred as opposed to hot because it drives structures to grow hierarchically, with small objects collapsing under their self-gravity first and merging

to form larger and more massive objects. Hierarchical structure formation is in very good agreement with observations of the cosmological evolution of structures. The ΛCDM model presents the cosmological constant in Einstein's equations as a natural candidate for dark energy. Finally, inflation is the most plausible mechanism for generating initial perturbations which correspond to the fluctuations in the matter density field and in the CMB.
In its most modern formulation, ΛCDM is parametrized by the following 6 parameters:

$$\{\rho_b, \rho_c, n_s, \mathcal{A}_s, \theta_\star, \tau\} \tag{1.5}$$

Here, ρ_b and ρ_c are the baryonic and dark matter densities respectively. The parameters n_s and $\mathcal{A}_s$, the spectral index and the scalar amplitude, fully define the inflationary fluctuations. Finally, $\theta_\star$ is the angular size of the CMB sound horizon and τ is the integrated optical depth to recombination.
The following table summarizes the key moments in the cosmological evolution of the Universe:

TABLE 1.1: The main events in the history of the Universe

Event	**Time**	**Energy**	**Temperature**	z
Quantum gravity	$< 10^{-43}$ s	10^{19} GeV	10^{32} K	?
Grand unification	$\sim 10^{-36}$ s	$\sim 10^{16}$ GeV	$\sim 10^{29}$ K	?
Inflation	$\gtrsim 10^{-34}$ s	$\lesssim 10^{15}$ GeV	$\lesssim 10^{28}$ K	?
Dark Matter decoupling	?	?	?	?
Baryogenesis	?	?	?	?
EW phase transition	10^{-10} s	100 GeV	10^{15} K	?
QCD phase transition	10^{-4} s	150 MeV	10^{12} K	?
Neutrino decoupling	1 s	1 MeV	10^{10} K	?
Nucleosynthesis	200 s	100 keV	10^9 K	10^8
Recombination	370,000 yrs	0.25 eV	2970 K	1100
First stars	10^8 yrs	4 meV	50 K	25-6
First galaxies	10^9 yrs	1.7 meV	20 K	1-7/8

Chapter 2

The Friedmann-Robertson-Walker Metric

2.1 Derivation

While a non-preturbative description of the evolution of the Universe at every scale seems like a hopeless task, the requirements of homogeneity and isotropy constrain the allowed geometries of the constant time hypersurfaces to the point that large scale properties can be treated analytically.
To exploit the properties of homogeneity and isotropy, one should however be mindful of the following caveat. Not all observers see the Universe as isotropic. For instance, observers that are locally boosted with respect to our own galaxy will see a dipole anisotropy in the Cosmic Microwave Background in the direction toward which they are moving. Therefore, when formulating the assumption of isotropy, one should specify that the validity is only restricted to a family of "typical" freely falling observers: those that move with the average velocity of typical galaxies in their respective neighborhoods.
Following this line of reasoning, it is useful to patch spacetime with a specific set of coordinates, called "comoving coordinates", which are attached to the motion of galaxies. These coordinates are constructed assuming that each galaxy carries with it a clock and is given a fixed set of spatial coordinates. The spacetime coordinates $(\mathbf{x}, t)$ of any event are therefore defined by taking $\mathbf{x}$ as the coordinate of the galaxy that is moving along where the event occurs, and by taking t as the time measured by that galaxy's clock. Thus, the coordinate system built this way is dragged along with the galaxies, which are in free fall.
The metric $g_{\mu\nu}$ in these coordinates will necessarily be characterized by very simple features, because of the conditions of isotropy and homogeneity. First, we note that clocks are in free fall, and therefore tell the proper time, so the proper time interval between two events at $(\mathbf{x}, t)$ and $(\mathbf{x}, t + dt)$ is:

$$-d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu = g_{00} dt^2 \tag{2.1}$$

and therefore this sets:

$$g_{00} = -1 \tag{2.2}$$

Next, the galaxy's trajectory of $\mathbf{x} =$ constant and $t = \tau$ satisfies the geodesics equation, because it is in free fall:

$$0 = \frac{dx^i}{d\tau^2} + \Gamma^i_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = \Gamma^i_{00} \tag{2.3}$$

Using (2.2) one then finds:

$$0 = g^{ij}\frac{\partial g_{j0}}{\partial t} \tag{2.4}$$

and because g^{ij} is non-singular, this equation implies:

$$0 = \partial_0 g_{j0} \tag{2.5}$$

In practice, the clocks attached to galaxies are set arbitrarily, so now suppose we would like to redefine the time on these clocks by a transformation $x^\mu \to x^{\mu'}$:

$$\begin{aligned} t' &= t + f(\mathbf{x}) \\ \mathbf{x}' &= \mathbf{x} \end{aligned} \tag{2.6}$$

The space-time metric element will transform as:

$$g'_{0i} = g_{0i} + \frac{\partial f}{\partial x^i} \tag{2.7}$$

There is a particular case where f can be chosen as to give $g'_{0i} = 0$, when the metric is manifestly spherically symmetric, such that the metric will take the general form:

$$d\tau^2 = dt^2 - D(r,t)dr^2 - 2E(r,t)drdt - F(r,t)r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.8}$$

As will be explained briefly, spherical symmetry is a valid supposition for the case of our Universe. The only non-vanishing component of g_{ti} is g_{tr} and, also, by virtue of (2.5), $E(t,r)$ is time-independent. At this point, the component g_{tr} can be eliminated by resetting the clocks with:

$$f(r) = -2\int^r E(r)dr \tag{2.9}$$

so that the metric (2.8) takes the form:

$$d\tau^2 = dt^2 - D(r,t)dr^2 - F(r,t)dr^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.10}$$

It can be shown that spaces that are isotropic and homogeneous possess the maximum number of Killing vectors, and are called for this reason "maximally symmetric spaces". The Universe's metric (2.10) is a case in which the whole spacetime is not maximally symmetric, but only its constant time hypersurfaces. It is a general result in differential geometry that the maximal symmetry of a certain subspace imposes very strong constraints on the metric of the whole space. A theorem states that if the whole spacetime has N dimensions, and its maximally symmetric subspace has M dimensions, it is always possible to choose the coordinates on the spatial hypersurfaces so that the metric is given by the following warp product:

$$-d\tau^2 = g_{ab}(v)dv^a dv^b + f(v)\tilde{g}_{ij}(u)du^i du^j \tag{2.11}$$

where $\tilde{g}_{ij}$ is the metric on the maximally symmetric subspace. In our case, $N = 4$ and $M = 3$, so there is one $v-$coordinates and three $u-$coordinates which, after being renamed accordingly, imposes a strict constraint on the form of (2.10), which becomes:

$$ds^2 = -dt^2 + a^2(t)\ {}^{(3)}ds^2 \tag{2.12}$$

where $a(t)$ is the scale factor of the Universe and:

$${}^{(3)}ds^2 = e^{2\beta(r)}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.13}$$

where a suitable rescaling of r was used in (2.10), after the time dependence of F and D was moved to a common factor $a(t)$. The fact that the functions $D(r,t)$ and $F(r,t)$ have a common time dependence which can be separated from the radial dependence should be intuitive, since a different functional form of the two functions on t would imply that observers located at different points in space measure contrasting time flows, which is prohibited by the homogeneity requirement (which, of course, we have yet to impose). The form of ${}^{(3)}ds^2$ is now relatively straightforward to deduce via symmetry arguments. In differential geometry, maximally symmetric spaces are uniquely specified by a "curvature constant" K and by the number of eigenvalues of the metric that are positve or negative. In the case at hand, one is looking for maximally symmetric 3-spaces with 3 positive eigenvalues, which, as is clear from (2.12), represent spatial hypersurfaces of spacetime. The choice of spherical symmetry in (2.8) is now understandable, since a maximally symmetric space will necessarily also be spherically symmetric. The Riemann and Ricci tensors of such spaces can be proved to satisfy the following relations:

$${}^{(3)}R_{ijkl} = K(g_{ik}g_{jl} - g_{il}g_{jk}) \tag{2.14}$$

$${}^{(3)}R_{jl} = 2Kg_{jl} \tag{2.15}$$

In differential geometry a space with these properties is called a space of constant curvature, given that a calculation of the Ricci scalar yields:

$${}^{(3)}R = 6K \tag{2.16}$$

The components of the Ricci tensor now can be calculated in a straightforward manner from (2.13), and they yield:

$$\begin{aligned} {}^{(3)}R_{11} &= \frac{2}{r}\partial_r\beta \\ {}^{(3)}R_{22} &= e^{-2\beta}(r\partial_r\beta - 1) + 1 \\ {}^{(3)}R_{33} &= (e^{-2\beta}(r\partial_r\beta - 1) + 1)\sin^2\theta \end{aligned} \tag{2.17}$$

Imposing these components to satisfy the relation (2.15), one finds the following form for $\beta(r)$:

$$\beta(r) = -\frac{1}{2}\log(1 - Kr^2) \tag{2.18}$$

Therefore, the metric (2.13) becomes:

$${}^{(3)}ds^2 = \frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.19}$$

The whole spacetime metric therefore takes the form:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-Kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right] \tag{2.20}$$

Finally, the curvature constant K can in principle take arbitrarily large positive or negative values, so a better variable k is usually preferred which is defined by the rescalings:

$$\begin{aligned} K \to k &\equiv \frac{K}{|K|} \\ r \to r' &\equiv \sqrt{|K|}r \\ a \to &\frac{a}{\sqrt{|K|}} \end{aligned} \tag{2.21}$$

which leave (2.20) invariant. Therefore, the only really relevant parameter is $K/|K|$ which can take the values $k = -1, 0, +1$. The metric finally takes the form:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right] \tag{2.22}$$

which is called the Friedmann-Robertson-Walker (FRW) metric, which is written in the coordinates used by galaxies in the course of their paths along the expansion of the Universe. This derivation, although technical, makes it clear that the main properties of spacetime can be captured with symmetry arguments.

Comoving Coordinates and Hubble's Law

In this section we would like to better clarify the notion of comoving observers and the physical meaning behind the coordinates (r, θ, ϕ) in the FRW metric. While this elucidation may seem tautological, given that the derivation of the metric started off by constructing the comoving coordinates themselves, it will be instrumental for the connection with Hubble's law and will serve as a useful summary.
First, one should be mindful that the convenience of a certain coordinate system does not mean that comoving observers are not moving away from one another. Consider for the time being a Universe with $k = 0$, which corresponds to flat space (see below), and two observers fixed at coordinates $(r_1, 0, 0)$ and $(r_2, 0, 0)$, with $r_1 < r_2$, where for convenience only a radial separation was chosen. At a specific instant in time, the proper distance L between the two objects will be:

$$L(t) = \int_{r_1}^{r_2} a(t)dr = a(t)\Delta r \tag{2.23}$$

where $\Delta r = r_2 - r_1$ is the coordinate distance. Since the Universe is expanding, the scale factor is such that $\dot{a} > 0$, so that the distance between to observers that are locally fixed at a certain coordinate is increasing. This is consistent with the procedure used to construct the metric, where the observers at a fixed $\mathbf{x}$ were in fact freely falling. Deriving (2.23) with respect to time, one finds Hubble's law:

$$\dot{L}(t) = H(t)L(t) \tag{2.24}$$

where the Hubble parameter was defined as:

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)} \tag{2.25}$$

At present time, $t = t_0$, the Hubble constant is denoted as $H(t_0) = H_0$, which is the parameter that was introduced in (1.1). However, with this version of Hubble's law, one should note the following. Firstly, the law in general needs to take into account "peculiar velocities" of galaxies, i.e. their local motion that is the result of local gravitational effects. To incorporate this effect, usually one lets Δr be a function of time, so that the full Hubble law reads:

$$\dot{L}(t) = H(t)L(t) + a(t)\dot{\Delta} r \tag{2.26}$$

Then, it is clear that the velocity of a certain observer has two contributions. The first is what is called historically the "Hubble flow", that is the motion of the observer, at fixed coordinates, resulting simply from the expansion of the Universe. The second contribution is its local coordinate motion due to possible gravitational effects in its neighborhood.
Finally, note that the style of Hubble's law appearing in (2.25) or (2.26) cannot technically be compared to Hubble's law as given by (1.1), since here the distance $L(t)$ is a proper distance, which is taken at a specific instant of time, and is therefore not an observable. Later, a more in depth clarification of Hubble's original distance-redshift relation will be given, where it will be found that more relevant measures of distance all approach the proper distance only when Δr is really small.
Essentially, the coordinate system (r, θ, ϕ) in the FRW metric is said to be "comoving" because an observer who is initially at rest at some coordinate, will remain at the said coordinate despite the action of a time varying gravitational field. That is the observers at $\mathbf{x} =$ constant evolve over time in a way that can be described as "attached" to the evolution of the geometry itself. Note that in the literature, coordinate distances are called "comoving distances", and proper distances are referred to as "physical distances". To switch between the two, one multiplies by the scale factor, as in (2.23):

$$\mathbf{x}_{\text{phys}} = a(t)\mathbf{x}_{\text{com}} \tag{2.27}$$

Physical coordinates are well suited for using the Newtonian intuition, as long as one limits the description of the dynamics within a certain length scale $\lambda = H^{-1}$. To see this more clearly, differentiating (2.27) one gets:

$$d\mathbf{x}_{\text{phys}} = a(d\mathbf{x}_{\text{com}} + H\mathbf{x}_{\text{com}}dt) \tag{2.28}$$

This equation can be inverted to find:

$$ad\mathbf{x}_{\text{com}} = d\mathbf{x}_{\text{phys}} - \mathbf{x}_{\text{phys}}Hdt \tag{2.29}$$

Using this coordinate change, the (flat) FRW metric becomes:

$$ds^2 = -2H\mathbf{x}_{\text{phys}} \cdot d\mathbf{x}_{\text{phys}}dt - (1 - H^2\mathbf{x}^2_{\text{phys}})dt + d\mathbf{x}^2_{\text{phys}} \tag{2.30}$$

Therefore, in physical coordinates, at distances such that $|\mathbf{x}_{\text{phys}}| \ll H^{-1}$, the FRW metric reduces to the Minkowski metric, where the Newtonian intuition holds. The

length scale H^{-1} is called the "Hubble horizon", a concept which will be examined in detail below.
We close out the discussion by offering two additional mathematical proofs that shed more light on the meaning of coordinates and comoving observers. First, consider an observer at a coordinate x^i that is initially at rest $\dot{x}^i = 0$. Its initial acceleration, provided by the geodesic equation, is given by:

$$\ddot{x}^i = -\Gamma^i_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = -\Gamma^i_{00}(\dot{x}^0)^2 \tag{2.31}$$

Given that the FRW metric implies $\Gamma^i_{00} = 0$, even the initial acceleration of the said observer will be null, so that he will remain fixed in his initial position.
Another approach consists in showing that the velocity fields tangent to the trajectories of static observers satisfy the geodesics equation. To verify this, consider the four velocity u^μ of a static observer in the FRW metric:

$$u^\mu = (-1, \mathbf{0}) \tag{2.32}$$

with the usual constraint $g_{\mu\nu}u^\mu u^\nu = -1$. The geodesics equation for the spatial components of u^μ is identically satisfied, while the one for u^0 reads:

$$\frac{du^0}{d\tau} = \frac{du^0}{dt}\frac{dt}{d\tau} = -\frac{1}{N^3}\frac{dN}{dt'} = -\Gamma^0_{00}(u^0)^2 \tag{2.33}$$

On the other hand:

$$\Gamma^0_{00} = \frac{1}{N}\frac{dN}{dt} \tag{2.34}$$

Therefore, even the component u^0 satisfies the geodesics equation.

Spatial Hypersurfaces

It is interesting to consider the geometrical properties of the spatial hypersurfaces, whose metric we recall here:

$${}^{(3)}ds^2 = \frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.35}$$

where the parameter k is related to the curvature constant K and can take the values $k \in \{-1, 0, +1\}$.
The case $k = 0$ corresponds to flat space, as can be readily verified by the following change of coordinates:

$$x_1 = r\sin\theta\cos\phi \quad x_2 = r\sin\theta\sin\phi \quad x_3 = r\cos\theta \tag{2.36}$$

under which the spatial element takes the usual form:

$${}^{(3)}ds^2 = dx_1^2 + dx_2^2 + dx_3^2 \quad k = 0 \tag{2.37}$$

This is an Euclidian 3-space denoted by $\mathbb{R}^3$, which has infinite extension since $-\infty < x_i < +\infty$ with $i = 1, 2, 3$. The case with $k = +1$ instead reads:

$${}^{(3)}ds^2 = \frac{dr^2}{1-r^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.38}$$

Transforming the r coordinate in the following way:

$$\arcsin r = \chi \tag{2.39}$$

leads to the metric of the 3-sphere S^3:

$${}^{(3)}ds^2 = d\chi^2 + \sin^2\chi[d\theta^2 + \sin^2\theta d\phi^2] \quad k = +1 \tag{2.40}$$

Here, as opposed to the flat space case, $a(t)$ has the interpretation of the radius of the Universe, since setting $k = +1$, t =const in (2.22) and calculating the volume leads to:

$$V_{S^3} = \int d^3x\, a(t)\sqrt{\det({}^{(3)}ds^2)} = 2\pi^2 a^3 \tag{2.41}$$

Therefore, the expansion or contraction of the spatial slices is equivalent to the expansion or contraction of the 3-dimensional surface S^3. In this case, since $\chi \in [0, 2\pi]$, the volume is finite but unbounded. For the flat model the scale factor cannot be regarded as any sort of physical radius, but only represents how the physical distance between comoving points scales as the space evolves.
Finally, the case of negative curvature $k = -1$ is the 3-hyperboloid $\mathbb{H}^3$ with coordinates:

$$\operatorname{arcsinh} r = \zeta \tag{2.42}$$

so that the 3-metric becomes:

$${}^{(3)}ds^2 = d\zeta^2 + \sinh^2\zeta[d\theta^2 + \sin^2\theta d\phi^2] \quad k = -1 \tag{2.43}$$

A three dimensional hyperbolic space has infinite volume, since $\zeta \in [0, +\infty]$.

New Coordinates and Conformal Time

It can be shown that the FRW metric is conformally flat in a suitable coordinate system. First, to get rid of non-trivial g_{rr} component, we introduce the coordinate χ using the same approach as above:

$$d\chi = \frac{dr}{\sqrt{1 - kr^2}} \tag{2.44}$$

Under this change of coordinates, the FRW metric becomes:

$$ds^2 = -c^2dt^2 + a^2(t)(d\chi^2 + S_k^2(\chi)d\Omega^2) \tag{2.45}$$

where:

$$S_k(\chi) \equiv \begin{cases} \sinh\chi & k = -1 \\ \chi & k = 0 \\ \sin\chi & k = +1 \end{cases} \tag{2.46}$$

Note how for a flat Universe ($k = 0$), there is no distinction between r and χ. A "conformal time" coordinate:

$$d\eta = \frac{dt}{a(t)} \tag{2.47}$$

finally accomplishes the goal of reducing the FRW metric to the Minkowski metric except for a conformal factor:

$$ds^2 = a^2(\eta)(-c^2 d\eta^2 + d\chi^2 + S_k^2(\chi) d\Omega^2) \tag{2.48}$$

An obvious application where the notion of conformal time results in a clear advantage over proper time t is spacetime diagrams. However, conformal time is a rather important concept in and of itself, for example when dealing with cosmological perturbations.

2.2 The motion of objects

The equations of motion for massive and massless point particles in the FRW metric are readily derived using the geodesics equation. It is convenient to write the metric as:

$$ds^2 = -dt^2 + g_{ij} dx^i dx^j = -dt^2 + a^2(t)\gamma_{ij} dx^i dx^j \tag{2.49}$$

The Christoffel's symbols then read:

$$\begin{aligned} \Gamma^{\mu}_{00} &= \Gamma^{0}_{\alpha 0} = 0 \\ \Gamma^{0}_{ij} &= a\dot{a}\gamma_{ij} \\ \Gamma^{i}_{0j} &= \frac{\dot{a}}{a}\delta^i_j \\ \Gamma^{i}_{jk} &= \frac{1}{2}\gamma^{il}(\partial_j \gamma_{kl} + \partial_k \gamma_{jl} - \partial_l \gamma_{jk}) \end{aligned} \tag{2.50}$$

Consider now the geodesics equation:

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma^{\mu}_{\alpha\beta} \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} = 0 \tag{2.51}$$

Using the four momentum of the point particle $P^\mu = m\, dx^\mu / d\tau$, and the fact that:

$$\frac{d}{d\tau} P^\mu(x^\alpha(\tau)) = \frac{dx^\alpha}{d\tau} \frac{\partial P^\mu}{\partial x^\alpha} = \frac{P^\mu}{m} \frac{\partial P^\mu}{\partial x^\alpha} \tag{2.52}$$

equation (2.51) becomes:

$$P^\alpha \frac{\partial P^\mu}{\partial x^\alpha} + \Gamma^{\mu}_{\alpha\beta} P^\alpha P^\beta = 0 \tag{2.53}$$

This is the form that is most convenient when deriving the equations of motion. However, a subtlety is that proper time is zero for massless particles, so another parameter is really needed in (2.53). Technically, nothing prevents from using a parameter λ to parametrise the path of a massless particle and, once chosen, there is only one continuation of it such that $x(\lambda)$ satisfies the geodesics equation, provided that $P^\mu \equiv dx^\mu / d\lambda$ is interpreted as the four momentum of the massless particle.

- Massive particles: consider the $\mu = 0$ component of (2.53):

$$P^0 \frac{dP^0}{dt} + \frac{\dot{a}}{a} g_{ij} P^i P^j = 0 \tag{2.54}$$

where isotropy arguments required that $\partial_i P^0$ vanish. We also introduce the "physical momentum" $p^2 \equiv g_{ij}P^iP^j$ which obeys the on-shell relation:

$$-(P^0)^2 + p^2 = -m^2 \tag{2.55}$$

Differentiating (2.55) and substituting it in (2.54) gives:

$$p\frac{dp}{dt} + \frac{\dot{a}}{a}p^2 = 0 \tag{2.56}$$

Finally, rearranging we have:

$$\frac{\dot{p}}{p} = -\frac{\dot{a}}{a} \tag{2.57}$$

This differential equation has a solution $p \sim a^{-1}$, meaning that the particles experience a drag caused by the expansion of the Universe. This also reveals why the comoving coordinates are a natural reference frame:

$$p = \frac{mv}{\sqrt{1-v^2}} \sim \frac{1}{a} \tag{2.58}$$

where $v^i = dx^i/dt$ is the peculiar velocity of the particle (remember that the peculiar velocity of an object is the velocity relative to the comoving frame) and $v^2 = g_{ij}v^iv^j$ as with the momentum. The first equality in (2.58) is just the usual special relativity relationship $dx^\mu/d\tau = u^\mu = (\gamma, \gamma v^i)$ with γ the local Lorentz factor. Now, consider a particle with initial physical (peculiar) velocity $v(t_1) \equiv v_1$. At a later time t_2 it will have a velocity:

$$v_2 = v_1\frac{a(t_1)}{a(t_2)} \tag{2.59}$$

The expansion of the Universe implies that $a(t_2) > a(t_1)$, and therefore $v_2 < v_1$, i.e. even if an observer has a nonzero initial velocity, he will come to rest in the comoving frame.

- Massless particles: take (2.54) with the new constraint $m = 0$:

$$-(P^0)^2 + p^2 = 0 \tag{2.60}$$

Using $P^0 = E$, one finds:

$$\frac{\dot{E}}{E} = \frac{\dot{a}}{a} \tag{2.61}$$

which implies that the energy scales as $E \sim a^{-1}$ for photons. The physical interpretation of such scaling is a stretch of the wavelengths proportional to the scale factor, giving rise to a decrease in energy. Interestingly, this is also the phenomenon at the base of cosmological redshift, which is the subject the next section

2.3 Cosmological redshift

The observation of light emitted by distant phenomena is an extremely effective approach to study of the properties of the Universe. Pioneering work, such as Hubble's finding of the redshifting of distant galaxies and the discovery of the CMB, as outlined in the introductory chapter, laid the foundations for the development of the Standard Model of Cosmology. The further analysis of these discoveries requires the knowledge of how light propagates as the Universe evolves in time. Above, photon wavelengths were found to get stretched as space expands, which is fundamentally the mechanism giving rise to redshifts of galaxy spectra. While equation (1.3) already introduced the nomenclature, this section will give more precise definitions. Consider the scenario in which a galaxy, positioned at (r_e, θ_e, ϕ_e) sends light signals to Earth, located at the origin of the coordinate system. Isotropy and homogeneity imply the freedom to choose the propagation of light in a straight line, i.e. $\theta = \phi =$ const along null geodesics $ds^2 = 0$. Using (2.48) with $k = 0$ for convenience yields:

$$0 = a(\eta)(-c^2 d\eta^2 + d\chi^2) \rightarrow \Delta\chi(\eta) = \pm c\Delta\eta \tag{2.62}$$

where the $\pm$ corresponds to outgoing and ingoing photons. If a wave crest is emitted at time t_e from the galaxy, the time t_0 when it reaches us is given by:

$$c^2(\eta(t_0) - \eta(t_e)) = \chi(r_e) - \chi(0) = \chi(r_e) \tag{2.63}$$

Since the comoving distance of the galaxy does not change with time, a successive wave crest emitted shortly after, at $t_e + \delta t_e$, reaches the origin at time $t_0 + \delta t_0$:

$$c^2(\eta(t_0 + \delta t_0) - \eta(t_e + \delta t_e)) = \chi(r_e) \tag{2.64}$$

Combining these two equations and noticing that for real applications we can approximate $\delta t_e \ll t_e$ and $\delta t_0 \ll t_0$, we find:

$$\frac{\delta t_0}{a(t_0)} = \frac{\delta t_e}{a(t_e)} \tag{2.65}$$

where we have used the definition of conformal time. Thus the period of the wave, hence its wavelength, increases proportionally to the scale factor:

$$\frac{\lambda_0}{\lambda_e} = \frac{\nu_e}{\nu_0} = \frac{\delta t_0}{\delta t_e} = \frac{1}{a(t_e)} \tag{2.66}$$

where we remember that we have the freedom to choose $a(t_0) = 1$. Clearly, this is the result found in (2.61), but now in a more concise way. Defining the relative change of wavelength as the redshift z, equation (1.3) is found:

$$a(t_e) = \frac{1}{1+z} \tag{2.67}$$

Naturally, the name "redshift" is a consequence of the fact that $\dot{a} > 0$, otherwise physical lengthscales would be contracting, giving rise to blueshifts. Interestingly,

this ties to the observations Hubble was carrying out in the 1920s. To see this, consider the expansion of the wave factor for nearby sources ($z \ll 1$) around $t = t_0$:

$$a(t_e) = 1 + (t_e - t_0)H_0 + ... \tag{2.68}$$

where $t_e - t_0$ is called the "look-back time". Using the definition of the redshift in (2.67), one then finds $z = H_0(t_0 - t_e) + ...$, but for close objects the simple relation $d/c = t_0 - t_e$ holds, where d is the proper distance (it is equal to L in (2.23)). Crucially, the non-relativistic limit of the formula for a longitudinal Doppler shift is:

$$z = \frac{\lambda_0}{\lambda_e} - 1 = \sqrt{\frac{1+\beta}{1-\beta}} - 1 = \sqrt{\frac{1+v/c}{1-v/c}} - 1 \approx \frac{v}{c} \tag{2.69}$$

so that one finds Hubble's law again:

$$v \approx cz \approx H_0 d \tag{2.70}$$

Basically, Hubble's discovery was of a cosmological redshift that masqueraded as a Doppler redshift, given his investigation of exclusively close-by sources. While the frequency shifts in the spectra of galaxies find a natural explanation in terms of Doppler shifts, as is clear from the derivation above, the cosmological redshift is of a different nature compared to a relativistic Doppler shift. Effectively, even galaxies that are moving with the Hubble flow, but have vanishing peculiar velocity, will experience a redshift. However, the two notions can be used interchangeably for close sources, which is what happened with the Hubble law. As was remarked above, all definition of distances converge for such sources. Instead, for objects that reside at very large distances, more attention needs to be given to a correct definition of distance.

2.4 Distances in cosmology

The measure of distances in cosmology is obviously fundamental for the understanding of the Universe's properties. There are, however, a lot of different effects to consider when trying to formulate rigorous definitions.
For starters, one should really be mindful that different techniques are used to measure distances of objects, depending if they are relatively close to Earth, say at redshifts not much greater than 0.1, or if they are found at cosmological distances. For close sources, say in our own Galaxy, a common method to measure distances is via the trigonometric parallax, which is the maximum angular radius π sweeped by the apparent position of any star due to the the annual motion of the Earth around the Sun. This notion of distance is referred to as d_{p} and is calculated as (stars in general satisfy $\pi \ll 1$):

$$d_{\mathrm{p}} = \frac{d_{\mathrm{E}}}{\pi} \tag{2.71}$$

where d_{E} is the mean distance of the Earth from the Sun, defined as the astronomical unit (AU). Another method consists in measuring the proper motion of a star $v_\perp$ perpendicular to the line of sight, that at a distance $d_\perp$ will appear to have moved

across the sky at a rate μ in radians/time:

$$d_\perp = \frac{v_\perp}{\mu} \tag{2.72}$$

However, the most common method of determining distances (of close sources) is based on the measurement of apparent luminosities of sources of known absolute luminosities. While the absolute luminosity L is the total energy emitted per second by the source, the apparent luminosity ℓ is the energy received by the observer per second per unit of receiving area. If the energy is emitted isotropically, in Euclidean geometry the relation between L and ℓ is found easily by imagining the object to be surrounded with a sphere whose radius is equal to the distance d_L between the object and Earth:

$$\ell = \frac{L}{4\pi d_\mathrm{L}^2} \tag{2.73}$$

Close by objects of known absolute luminosities are, for example, main sequence stars, red clump stars and Cepheids variables. When these indicators are not bright enough, if the observations probe farther distances, one can use whole galaxies, or Type Ia supernovae as reference sources. For galaxies, some interesting phenomenological relations between their properties and their absolute luminosities have been encapsulated by the Tully-Fisher relation, the Faber-Jackson relation and the fundamental plane relation.
For measurements of distances at large redshifts, say $z > 0.1$, the effects of cosmological expansion cannot be neglected.

Luminosity distance

The method used to determine d_L with (2.73) is powerful enough that it is extendable to higher redshifts. Consider a source at redshift z, with comoving distance being:

$$\chi(z) = \int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_0^z \frac{dz}{H(z)} \tag{2.74}$$

While (2.73) was derived under the assumption of a static Euclidian space, in a more complex expanding spacetime, the equation is modified in three fundamental ways:

- First, the radius of the sphere is not χ, but rather $S_k(\chi)$, as laid out by the FRW metric. The sphere surface is then $4\pi S_k(\chi)^2$. Furthermore, when light reaches Earth at time t_0, the sphere has an actual area stretched by $a^2(t_0)$, because of expansion. In total, the denominator of (2.73) becomes:

$$4\pi a^2(t_0) S_k(\chi)^2 \tag{2.75}$$

- The photons reach Earth with a decreased rate, again because of expansion. This effect reduces the apparent luminosity by a factor of $a(t_1)/a(t_0) = 1/(1+z)$.

- The observed photons lose an energy proportional to $1/(1+z)$ during their trip to Earth, resulting in a net decrease in ℓ once again.

For these reasons, the correct formula relating the apparent and the intrinsic luminosity of a source at redshift z is:

$$F = \frac{L}{4\pi S_k(\chi)^2(1+z)^2} \equiv \frac{L}{4\pi d_{\mathrm{L}}^2(z)} \tag{2.76}$$

where, with a slight abuse of notation, we have defined the "luminosity distance" as:

$$d_{\mathrm{L}}(z) = (1+z)S_k(\chi(z)) \tag{2.77}$$

An expression for $d_{\mathrm{L}}(z)$ is usually given as a power series in z, although in some cases it is possible to find a closed analytic form. Expanding first $a(t)$ for $z \ll 1$ one gets:

$$a(t) = 1 + H_0(t-t_0) - \frac{1}{2}q_0 H_0^2(t-t_0)^2 + ... \tag{2.78}$$

where the "deceleration parameter" at present time q_0 was defined to be:

$$q_0 \equiv -\frac{\ddot{a}}{aH^2}\Big|_{t_0} \tag{2.79}$$

Given the definition of cosmological redshift, one finds:

$$z = \frac{1}{a(t_1)} - 1 = H_0(t_0 - t_1) + \frac{1}{2}(2+q_0)H_0^2(t_0-t_1)^2 + ... \tag{2.80}$$

which can be inverted to give:

$$H_0(t_0 - t_1) = z - \frac{1}{2}(2+q_0)z^2 + ... \tag{2.81}$$

The comoving distance can then be rewritten by making use of (2.78) and (2.81):

$$\begin{aligned}\chi(z) = c\int_{t_1}^{t_0}\frac{dt}{a(t)} &= c\int_{t_1}^{t_0} dt(1 - H_0(t-t_0) + ...) \\ &= c(t_0 - t_1) + \frac{1}{2}\frac{H_0}{c}c^2(t_1-t_0)^2 + ... \\ &= \frac{c}{H_0}\left(z - \frac{1}{2}(1+q_0)z^2 + ...\right)\end{aligned} \tag{2.82}$$

If the geometry of the Universe is known, computing $S_k(\chi(z))$ with (2.82) yields an estimation of the luminosity distance of a source at redshift z, i.e. the redshift-distance relation. For example, if the Universe is spatially flat $k = 0$, then $S_k(\chi) = \chi$, and:

$$d_{\mathrm{L}}(z) = \frac{c}{H_0}\left(z + \frac{1}{2}(1-q_0)z^2 + ...\right) \tag{2.83}$$

Immediately one notices the first order approximation $d_{\mathrm{L}}(z) \simeq cz/H_0$, which is exactly Hubble's law. In general, however, the Hubble's law will be true, at first order, for all geometries, since the curvature of space will not be a factor for nearby sources, i.e. for $z \ll 1$.

Angular diameter distance

An alternative way to measure the distance of a given distant source is by measuring its angular amplitude. Given a source at comoving distance $\chi(z)$ which has emitted a photon at time t_1, and with transverse size D, its angular size is, in a static Euclidian space:

$$\delta\theta = \frac{D}{\chi} \tag{2.84}$$

where we assumed $\delta\theta \ll 1$, which is satisfied by very distant objects to a high degree of precision. If light travels radially, then from the FRW metric it follows that the angular diameter of the source is:

$$\delta\theta = \frac{D}{a(t_1)S_k(\chi)} \equiv \frac{D}{d_{\rm A}(z)} \tag{2.85}$$

where we have defined the "angular diameter distance":

$$d_{\rm A}(z) = \frac{S_k(\chi)}{1+z} \tag{2.86}$$

Notice the difference with the definition of the luminosity distance (2.77). Whereas for $d_{\rm L}$ we were interested in the moment of observation $a(t_0)$, we are now taking the time of emission $a(t_1)$. Furthermore, a perturbative expression for $d_{\rm A}(z)$ is readily achieved just by virtue of its relation to $d_{\rm L}(z)$:

$$d_{\rm A} = \frac{d_{\rm L}}{(1+z)^2} \tag{2.87}$$

Therefore:

$$d_{\rm A}(z) = \frac{c}{H_0}\left(z - \frac{1}{2}(1+q_0)z^2 + ...\right) \tag{2.88}$$

Notice how, for really close sources, the first order term gives $d_{\rm L} = d_{\rm A} \simeq cz/H_0$, i.e. Hubble's law. This, as remarked above, is true because spacetime is locally flat, and in Euclidian static space we know that there is only one distance, as corroborated by our everyday experience. It can also be proved that even the distances d_P and $d_\perp$, as defined by (2.71) and (2.72), all converge when $z \ll 1$:

$$d_{\rm A} \simeq d_{\rm L} \simeq d_\perp \simeq d_{\rm P} \qquad z \ll 1 \tag{2.89}$$

The distinction among different measures of distance only becomes important for objects at cosmological distances.

2.5 Horizons in the Universe

Due to the speed of light being finite, there are regions of the Universe that are and will be unaccessible to us, even if we wait an infinite amount of time. In General Relativity, these regions are defined by horizons and define the causal structure of the Universe.

Particle Horizon

The first of these horizons is what is called the "particle horizon" or the "cosmological horizon". It is the maximum distance light could have traveled from the "beginning" of the Universe, so that it defines the observable Universe. To be precise, the following question captures the essence of the particle horizon: given a comoving observer at coordinates $(\chi_0, \theta_0, \phi_0)$, for what values of (χ, θ, ϕ) would a light signal emitted at time $t = 0$ (the "beginning" of the Universe) reach the observer at time t? Given the FRW metric, the answer is rather straightforward. First of all, let's place the receiving observer at a convenient point in space by setting $\chi_0 = 0$. Then, a light signal satisfies the geodesics equation with $ds^2 = 0$ and $d\phi = d\theta = 0$ (because light moves on great circles). If this light signal reaches the origin at time t, then the comoving distance from the observer is given by:

$$d_{\text{h,comoving}}(\eta) = \chi = \eta = \int_0^t \frac{dt'}{a(t')} \tag{2.90}$$

In other words, the value of conformal time now η_0 measures the extent of the observable Universe at present time. The proper particle horizon is defined by the usual correspondence between comoving and proper distances:

$$d_{\text{h}}(\eta) = a(\eta) \int_0^t \frac{dt}{a(t)} \tag{2.91}$$

Notice that the implicit assumption of $t = 0$ corresponding to the beginning of the Universe is not entirely justified, given our present ignorance about the state of the Universe beyond the era of BBN, at 1 MeV. Again, extrapolating the theoretical picture to 100 GeV, the equation (2.90) should really be written as:

$$\eta = \int_0^{t_{100\,\text{Gev}}} \frac{dt'}{a(t')} + \int_{t_{100\,\text{Gev}}}^t \frac{dt'}{a(t')} \tag{2.92}$$

Because the Universe may have gotten to truly extreme conditions in the far past, there is no reason to trust classical GR at such high energies (the first integral), so that the definition of "observable Universe" may make not much sense at all. In fact, the discussion of inflation will make clear that the name "observable Universe" itself is usually meant to be the distance light could have traveled from some time after inflation ended, where extrapolating the classical theory is still sensible.

Event Horizon

Another type of horizon is the event horizon, which differs from the particle horizon in that it is the maximum distance light, emitted at given time t, could travel to the infinite future. Again, the FRW metric allows us to easily calculate this distance, proceeding in the following way. Given a comoving observer at coordinates $(0, \theta_0, \phi_0)$ (again, homogeneity permits to set $\chi_0 = 0$) that emits a light signal at time t, the distance it travels up to $t = \infty$ is:

$$d_{\text{e,comoving}}(\eta) = \int_\eta^{\eta_f} d\eta = \eta_f - \eta = \int_t^\infty \frac{dt'}{a(t')} \tag{2.93}$$

with its proper counterpart being:

$$d_{\mathrm{e}}(\eta) = a(\eta) \int_t^{\infty} \frac{dt'}{a(t')} \tag{2.94}$$

After a moment's thought, one notices that this definition really is just the "reverse" of the particle horizon, in the sense that if the beginning and end of the Universe were reversed, the two concepts would switch. This is shown more clearly in the following figure, which depicts the two horizons:

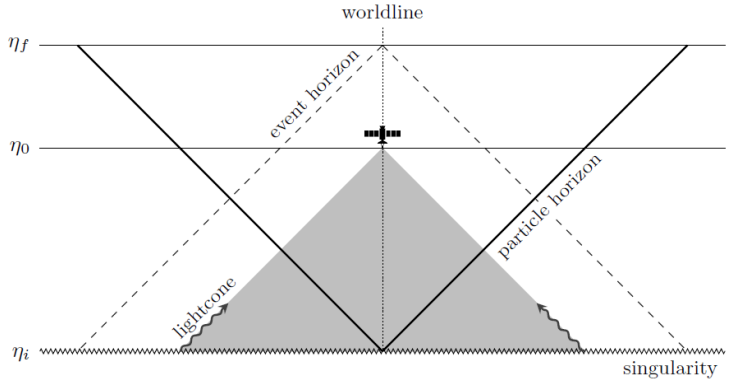

FIGURE 2.1: The particle and the event horizon, in conformal time.

For the moment the identities $\eta_i = \eta(t = 0) = 0$ and $\eta_f = \eta(t = \infty) \neq \infty$ (infinite time does not mean infinite conformal time) were only postulated, but will be derived later on, where the dynamics of the Universe are discussed in detail. Moreover, the specularity of the definitions for $d_e(\eta)$ and $d_h(\eta)$ implies that just as the particle horizon is sensitive to the initial conditions (the integral (2.91) goes down to $t = 0$), the event horizon is sensitive to the final conditions (the integral (2.93) goes until $t = \infty$). Practically, this is because there is still a great deal of uncertainty regarding the far future of the Universe, this time not because of problems intrinsic to GR, but because the entity that controls future evolution is dark energy, whose nature is still really unresolved.

Hubble Horizon

A final important concept is that of the "Hubble horizon". Although the name suggests otherwise, this quantity does not represent an horizon as it is usually intended in GR. As we have established, Hubble's law dictates that the velocity of galaxies on the Hubble flow is proportional to their proper distance:

$$v = H(t)d \tag{2.95}$$

Replacing v with the speed of light c gives the proper distance d_{HR} above which observers recede with a speed that's greater than that of light:

$$d_{\mathrm{HR}} = \frac{1}{H(t)} \tag{2.96}$$

The comoving Hubble horizon is instead:

$$d_{\mathrm{HR}} = \frac{1}{a(t)H(t)} \tag{2.97}$$

This is not in contrast with Special Relativity, as this velocity is not measured in any inertial frame. No observer is overtaking a light beam and locally, in the galaxy's position, observers measure the speed of light as c. Taking $(d_{HR})^3$, one gets what is usually called the "Hubble volume" or the "Hubble sphere". On the one hand, the Hubble sphere is not a measure of causality, because if two objects are separated by a distance greater than a Hubble length, it would be still possible for them to communicate. In fact, this is the case if the Universe is expanding and $\ddot{a} < 0$, i.e it is also decelerating, meaning that the Hubble sphere is actually increasing. After some time, the Hubble sphere will catch up with the light ray, which in turn will enter a zone where space is expanding slower than the speed of light, making it possible to freely flow to destination. If instead the Universe was expanding in an accelerating fashion, the Hubble sphere would decrease, and two objects separated by more than a Hubble length would forever be out of causal contact. On the other hand, this discussion implies that the Hubble horizon has got at least something to do with causality, despite not really being an horizon. Indeed, one finds that it is related to the particle horizon by rearranging the integral (2.90) starting at some time t_i:

$$d_h(\eta) = \int_{t_i}^{t} \frac{dt}{a} = \int_{a_i}^{a} \frac{da}{a\dot{a}} = \int_{\log a_i}^{\log a} (aH)^{-1} d\log a \tag{2.98}$$

Basically, the particle horizon is the logarithmic integral of the comoving Hubble horizon.

Chapter 3

The dynamics of expansion

Up until now the results obtained were all very general, that is they were independent of the underlying assumptions regarding the dynamics of expansion. Symmetry arguments were essentially the only tools that allowed us to determine the metric:

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right) \\ &= ds^2 = a^2(\eta)(-d\eta^2 + (d\chi^2 + S_k^2(\chi)d\Omega^2)) \end{aligned} \tag{3.1}$$

The dynamics of the expansion, encapsulated by the scale factor $a(t)$, are governed by Einstein field equations:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{3.2}$$

3.1 The Einstein Equations

The field equations (3.2), needed to determine the scale factor, are composed of a curvature part on the left hand side, and a part regarding the matter-energy content. Both need to be determined.

Curvature

By virtue of the components of the affine connection calculated in (2.50), the only non-vanishing components of the Ricci tensor are:

$$\begin{aligned} R_{ij} &= \frac{\partial \Gamma^k_{ki}}{\partial x^j} - \left(\frac{\partial \Gamma^k_{ij}}{\partial x^k} + \frac{\partial \Gamma^0_{ij}}{\partial t}\right) + (\Gamma^0_{ik}\Gamma^k_{j0} + \Gamma^k_{i0}\Gamma^0_{jk} + \Gamma^l_{ik}\Gamma^k_{jl}) \\ &\quad - (\Gamma^k_{ij}\Gamma^l_{kl} + \Gamma^0_{ij}\Gamma^l_{0l}) \\ R_{00} &= \frac{\partial \Gamma^i_{i0}}{\partial t} + \Gamma^i_{0j}\Gamma^j_{0i} \end{aligned} \tag{3.3}$$

The derivatives of the Christoffel's and the multiplicative terms read:

$$\begin{aligned} \frac{\partial \Gamma^0_{ij}}{\partial t} &= \gamma_{ij}\frac{d}{dt}(a\dot{a}) \quad \Gamma^0_{ik}\Gamma^k_{i0} = \gamma_{ij}\dot{a}^2 \quad \Gamma^0_{ij}\Gamma^l_{0l} = 3\gamma_{ij}\dot{a}^2 \\ \frac{\partial \Gamma^i_{i0}}{\partial t} &= 3\frac{d}{dt}\left(\frac{\dot{a}}{a}\right) \quad \Gamma^i_{0j}\Gamma^j_{i0} = 3\left(\frac{\dot{a}}{a}\right)^2 \end{aligned} \tag{3.4}$$

Using these relations in (3.3) yields:

$$\begin{aligned} R_{00} &= -3\frac{\ddot{a}}{a} \\ R_{ij} &= \left(\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{kc^2}{a^2}\right) \end{aligned} \tag{3.5}$$

Given the Ricci tensor, it is straightforward to compute the Ricci scalar:

$$R = 6\left(\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right) \tag{3.6}$$

Finally, we state the components of the Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - 1/2Rg_{\mu\nu}$:

$$\begin{aligned} G^0_0 &= -3\left(\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right) \\ G^i_j &= -\left(2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right)\delta^i_j \end{aligned} \tag{3.7}$$

The energy-momentum tensor

To complete the Einstein equations, the form of the energy-momentum tensor $T_{\mu\nu}$ for all the fields (matter, radiation, dark energy) present in the Universe needs to be specified. Just as the form of the metric $g_{\mu\nu}$ is dictated by symmetry arguments, differential geometry imposes strict constraints on the form of any tensor that is defined on a spacetime with a maximally symmetric subspace, such as the FRW spacetime. Intuitively, however, one is only interested in the smooth property of matter, since our description is still limited to large-scale properties, and isotropy imposes no mean net momentum or velocity of the fluid. Therefore, all that is left are the mean density and pressure as the only properties that characterize the matter-energy content. The only form of the energy-momentum tensor that is allowed is necessarily very simple:

$$T^\mu_\nu(x) = \begin{pmatrix} -\rho(t) & & & \\ & P(t) & & \\ & & P(t) & \\ & & & P(t) \end{pmatrix} \tag{3.8}$$

Where the requirement of homogeneity took out the $\vec{x}$ dependence from the pressure and density fields. Clearly, symmetry arguments show that the energy-momentum

tensor of the Universe necessarily takes the same form as for a perfect fluid. The explicitly covariant form of $T_{\mu\nu}$ reads:

$$T_{\mu\nu} = (\rho + P)\, u_\mu u_\nu + P g_{\mu\nu} \tag{3.9}$$

where u^μ is the relative 4-velocity between the fluid and a comoving observer, i.e. the fluid's peculiar velocity. Notice that, in the comoving frame, with $u^\mu = (1, 0, 0, 0)$, the covariant expression yields (3.8). This was expected since, as argued for the metric, isotropy and homogeneity only hold for the class of observers that are comoving. The energy-momentum tensor obeys the conservation equation as a consequence of diffeomorphism invariance of Einstein's theory:

$$\nabla_\mu T^\mu_\nu = 0 \tag{3.10}$$

Expanding it:

$$\nabla_\mu T^\mu_\nu = \partial_\mu T^\mu_\nu + \Gamma^\mu_{\mu\lambda} T^\lambda_\nu - \Gamma^\lambda_{\mu\nu} T^\mu_\lambda = 0 \tag{3.11}$$

These are four separate equations, one for every value of ν. The spatial components:

$$\nabla_\mu T^\mu_i = 0 \tag{3.12}$$

are identically satisfied. The time component of the conservation equation is instead more interesting:

$$\partial_\mu T^\mu_0 + \Gamma^\mu_{\mu\lambda} T^\lambda_0 - \Gamma^\lambda_{\mu 0} T^\mu_\lambda = 0 \tag{3.13}$$

This reduces to:

$$\dot\rho + 3\frac{\dot a}{a}(\rho + P) = \dot\rho + 3H(\rho + P) = 0 \tag{3.14}$$

This equation is the continuity equation in an expanding spacetime. In order to solve it, one needs an equation of state which relates pressure and density:

$$P = P(\rho) \tag{3.15}$$

In general, most cosmological fluids are well described by a constant equation of state:

$$P = w\rho \tag{3.16}$$

With this in mind, the continuity equation reads:

$$\frac{\dot\rho}{\rho} = -3(1 + w)\frac{\dot a}{a} \tag{3.17}$$

the solution of which is readily obtained:

$$\rho \propto a^{-3(1+w)} \tag{3.18}$$

In other words, the dilution of energy of a fluid in an expanding spacetime depends on the equation of state of the fluid itself.

Equations of state

The equations of state for different fluids are derived using statistical mechanics. Consider a gas of particles that are at thermodynamic equilibrium with temperature T, inside a box of volume V. Then, the number density of particles with momentum in the interval $(p, p+dp)$ is given by:

$$dn = \frac{g}{(2\pi\hbar)^3} f(p,T) 4\pi p^2 dp \tag{3.19}$$

where $f(p,T)$ is the distribution function and the g factor is the spin degeneracy of the particles. The thermodynamic variables are given by:

$$\begin{aligned} \frac{N}{V} &= \int_0^\infty dn \\ \rho = \frac{E}{V} &= \int_0^\infty dn\, E \\ \frac{P}{V} &= \frac{1}{3}\int_0^\infty dn\, pv = \frac{1}{3}\frac{\int_0^\infty dn\, pv}{\int_0^\infty dn} n = \frac{1}{3} n\langle pv\rangle \end{aligned} \tag{3.20}$$

where $E(p) = \sqrt{m^2+p^2}$ is the particles' energy. All that's left to do is to evaluate P for every particle species in the Universe. Starting with non-relativistic matter, one has:

$$E(p) = \sqrt{m^2+p^2} \simeq mc^2 \qquad v = \frac{p}{m} \tag{3.21}$$

This is the non-relativistic limit, and it applies to both dark energy and baryonic matter. The pressure then is:

$$P = \frac{1}{3} n\langle pv\rangle = \frac{1}{3} nm\langle v^2\rangle = \frac{1}{3} nE\langle v^2\rangle \simeq 0 \tag{3.22}$$

where the last approximation is due to the non-relativistic nature of the particles. For this reason:

$$P_{\text{matter}} = 0 \tag{3.23}$$

which implies $w_{\text{matter}} = 0$. The dilution of the energy density is then, from (3.18):

$$\rho_{\text{matter}} \propto a^{-3} \tag{3.24}$$

That is, the matter energy density scales as the inverse of the volume, with size a^{-3}. For radiation (by radiation we mean any ultra-relativistic species. At present time the only ones are photons and neutrinos, but every particle was once ultra-relativistic, as the Universe was very hot) the following properties hold:

$$E(p) = \sqrt{m^2+p^2} \simeq p \qquad v(p) = 1 \tag{3.25}$$

The pressure then becomes:

$$P = \frac{1}{3} n\langle pv\rangle = \frac{1}{3} n\langle p\rangle \tag{3.26}$$

whereas the energy density:

$$\rho = \frac{E}{V} = \int_0^\infty dn\ p = \frac{\int_0^\infty dn\ pc}{\int_0^\infty dn}\ n = n\langle p \rangle \tag{3.27}$$

Putting (3.27) and (3.26) together:

$$P_{\text{radiation}} = \frac{1}{3}\rho \tag{3.28}$$

which means $w_{\text{radiation}} = 1/3$, leading to an energy dilution of the form:

$$\rho_{\text{radiation}} \propto a^{-4} \tag{3.29}$$

Again, the energy density scales as the volume a^{-3}, but this time a factor of a^{-1} is added because radiation is redshifted.
The last component that remains to be analysed is dark energy, for which the current level of knowledge relies on the following equation of state:

$$P_\Lambda \simeq -\rho \tag{3.30}$$

that is $w_\Lambda = -1$, so that its energy density stays constant:

$$\rho_\Lambda \propto a^0 \tag{3.31}$$

Since the Universe is expanding, energy has to be created continuously to sustain the expansion. A natural candidate that satisfies this behavior is vacuum energy because, as the Universe expands, more space is being created and therefore the energy increases in proportion to the volume, leading to a constant energy density. General Relativity predicts the existence of such a fluid via the cosmological constant. In fact, the Einstein field equations allow for the inclusion of the cosmological constant into the energy-momentum tensor:

$$G_{\mu\nu} = 8\pi G(T_{\mu\nu} + T^\Lambda_{\mu\nu}) \tag{3.32}$$

where:

$$T^\Lambda_{\mu\nu} = -\frac{\Lambda}{8\pi G} g_{\mu\nu} = -\rho_\Lambda g_{\mu\nu} \tag{3.33}$$

Comparing this form with (3.9) implies $P = -\rho$. In general, the terms "vacuum energy" and "dark energy" are used interchangeably, to mean a certain entity that satisfies an equation of state such that $w = -1$.

3.2 The Friedmann Equations

All is now set to solve the Einstein equations. Below, we summarise the main components, starting with the non-zero components of the Einstein tensor $G_{\mu\nu}$:

$$\begin{aligned} G_0^0 &= -3\left(\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right) \\ G_j^i &= -\left(2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right)\delta_j^i \end{aligned} \tag{3.34}$$

Then, the form of the energy-momentum tensor is that of a perfect fluid, which in the comoving frame reads:

$$T_\nu^\mu(x) = \begin{pmatrix} -\rho(t) & & & \\ & P(t) & & \\ & & P(t) & \\ & & & P(t) \end{pmatrix} \tag{3.35}$$

which has to satisfy the conservation equation, giving a scaling of the energy density of:

$$\rho \propto a^{-3(1+w)} \tag{3.36}$$

The total energy density is the sum of the energy density of every component in the Universe:

$$\rho = \rho_{\text{matter}} + \rho_{\text{radiation}} + \rho_\Lambda \tag{3.37}$$

which scale as:

$$\rho_{\text{matter}} \propto a^{-3} \quad \rho_{\text{radiation}} \propto a^{-4} \quad \rho_\Lambda \propto a^0 \tag{3.38}$$

There are only two components of the Einstein equations, namely the 00 component, and any one of the spatial components ij, since G_j^i is proportional to δ_j^i, as is dictated by the requirement of isotropy. The 00 equation reads:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{3.39}$$

This equation is called the "Friedmann equation" and is particularly important. Given initial conditions, from this equation it is rather straightforward to calculate $a(t)$, which controls the dynamics of the whole Universe. The ρ present in this equation is to be understood as a sum, namely the one found in (3.37). The spatial part of the Einstein equations yields the (second) Friedmann equation:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \tag{3.40}$$

The two equations (3.39) and (3.40) control the large-scale dynamics of our Universe, in the sense that they can be solved for the functional form of the scale factor. Before moving on to solve the equations themselves, it is interesting to see how a multitude of useful insights can be gained through some simple considerations that don't involve specific solutions.

The Newtonian perspective and the fate of the Universe

The main intuition behind the Friedmann equations is that they are simply a statement about conservation of energy in the Newtonian framework. To see this, consider a non-relativistic system containing a distribution of matter ($P = 0$) with density $\rho(t)$ and organized in a spherical fashion, where the sphere has radius $a(t)$. If a test particle is placed on the surface of the said sphere, conservation of energy states that:

$$\frac{1}{2}m\dot{a}^2 - G\frac{mM}{a} = E \tag{3.41}$$

where E is the total energy and $M = 4/3\pi a^3(t)\rho(t)$ is the mass of the matter distribution. Rewriting this equation leads to the Friedmann form:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) + \frac{2E}{ma^2} \tag{3.42}$$

Clearly, this matches the the first Friedmann equation (3.39), although on two conditions. First, the density term needs to accommodate both radiation and dark energy, so that ρ becomes the sum of all these terms. Secondly, the identification $2E/m$ with -2κ needs to be made. This means that the total energy of the system is related to the spatial curvature k. Since the energy is eventually related to escape velocity, this implies that the curvature actually determines the fate of the Universe. If, for example, the energy of the system is 0, or equivalently the space is flat, then the sphere will expand forever, with a velocity that tends to 0 (note that, just because we are talking about a spherical matter distribution, it doesn't have any implication for the spatial geometry, which can be flat, open or closed). On the other hand, one can reason in the opposite way, because it is the density $\rho(t)$ that specifies the curvature of space in the first place through the Einstein equations, so in truth the two concepts are tightly linked.
To make even more explicitly the connection between the two, curvature and energy, we note that a flat Universe today ($t = t_0$) corresponds to the following "critical density":

$$\rho_{\text{crit},0} = \frac{3H_0^2}{8\pi G} \simeq 1.26 \times 10^{11} M_\odot \text{Mpc}^{-3} \tag{3.43}$$

Now, in Cosmology it is convenient to scale all the densities to the critical density, and work with the adimensional parameters Ω_i defined as:

$$\Omega_i = \frac{\rho_{i,0}}{\rho_{\text{crit},0}} \quad i \in \{\text{Matter, Radiation, Dark Energy}\} \tag{3.44}$$

Using these new parameters, the first Friedmann equation can be rewritten as:

$$\frac{H^2}{H_0^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda \tag{3.45}$$

where, as done in the literature, the adimensional energy density parameter associated with curvature is defined as $\Omega_k = -kc^2/H_0^2$. Evaluating (3.45) at present time yields an important constraint:

$$1 = \sum_i \Omega_i \equiv \Omega_0 + \Omega_k \tag{3.46}$$

Where the total energy content was defined as $\Omega_0 \equiv \Omega_r + \Omega_m + \Omega_\Lambda$. This implies that the sign of $\Omega_0 - 1$ is related to that of Ω_k:

$$\Omega_0 - 1 = -\Omega_k = \frac{kc^2}{H_0^2} \tag{3.47}$$

In other words, if the energy content of the Universe happens to satisfy $\Omega_0 - 1 = 0$, the spatial curvature will be zero, and the expansion will continue forever; this is the case of a flat Universe. The same line of reasoning can be carried out for the other cases, $\Omega - 1 < 0$ or $\Omega - 1 > 0$, which are the cases of open and closed Universes for which, respectively, the expansion will continue again forever or halt:

$$\begin{aligned} k = +1 &\longleftrightarrow \Omega_0 > 1 \quad \text{Closed Universe} \\ k = 0 &\longleftrightarrow \Omega_0 = 1 \quad \text{Flat Universe} \\ k = -1 &\longleftrightarrow \Omega_0 < 1 \quad \text{Open Universe} \end{aligned} \tag{3.48}$$

Notice one final important consequence of the last form of the Friedmann equation (3.45). The different scalings of the various energy densities imply that the Universe underwent different phases in which only one of the species dominated. In particular, for very early times $a \ll 1$, and the Universe underwent a period of radiation domination (RD), whereas only later and until close to present day was the evolution controlled by matter, in the period of matter domination (MD). Finally, only in the far future (actually, even today. See below) the dynamics are specified by dark energy in the period of dark energy domination (ΛD). This last period is still subject to uncertainty, as was outlined in the discussion of the cosmological event horizon. Since the nature of dark energy is still subject to intense study and research, it is not known if, for instance, its equation of state is really a constant $w_\Lambda = -1$ or if it depends on time $w_\Lambda = w_\Lambda(a)$. In this last case, the dark energy domination period will need to be revised.
In mathematical terms, the following limits of (3.45) hold:

$$\begin{aligned} \frac{H^2}{H_0^2} &\approx \Omega_r a^{-4} \longleftrightarrow \text{Early Universe (RD)} \\ \frac{H^2}{H_0^2} &\approx \Omega_m a^{-3} \longleftrightarrow \text{Late Universe (MD)} \\ \frac{H^2}{H_0^2} &\approx \Omega_\Lambda \longleftrightarrow \text{Today and the Future } (\Lambda\text{D}) \end{aligned} \tag{3.49}$$

Acceleration or deceleration?

During the study of distances in cosmology, we introduced the deceleration parameter q_0:

$$q_0 = -\frac{\ddot{a}}{aH^2}\Big|_{t_0} \tag{3.50}$$

Not surprisingly, the Friedmann equations allow us to find an analytical expression for this parameter. Using the second Friedmann equation (3.40) and the definition

of critical density (3.43), one finds:

$$q_0 = \frac{1}{2}\sum_i \Omega_i(1+3w_i) \tag{3.51}$$

where, as in (3.44), the index i spans all species. Now, using the various equations of state, this equation simplifies to:

$$q_0 = \frac{1}{2}(\Omega_m + 2\Omega_r - 2\Omega_\Lambda) \tag{3.52}$$

Therefore, the adimensional density parameters Ω_i, determine if the Universe today is accelerating $q_0 < 0$ or if it's decelerating $q_0 > 0$. Obviously, this calculation can be done for every instant of cosmic time, providing even more insight into the dynamics of the Universe.

The age of the Universe

A final, useful application of the Friedmann equations is the calculation of the age of the Universe. To estimate this quantity, the following relationship can in fact be used:

$$\frac{da}{dt}\frac{1}{a} = H \tag{3.53}$$

This implies:

$$\int_0^{t_0} dt = \int_0^1 \frac{da}{a}\frac{1}{H} \tag{3.54}$$

where, again, the extrapolation to the past is intended to be valid up until $t_{100\,\mathrm{GeV}}$ or some time after inflation, but surely not $t = 0$. For the time being, however, we keep the lower bound as it is and calculate the age of the Universe today as:

$$t_0 = \frac{1}{H_0}\int_0^1 da\,\left[\Omega_m a^{-1} + \Omega_r a^{-2} + \Omega_\Lambda a^{-2} + \Omega_k\right]^{-1/2} \tag{3.55}$$

The major takeaway from these properties derived from the Friedmann equations is that the measurement of the adimensional density parameters Ω_i is of fundamental importance. These parameters specify uniquely the scale factor $a(t)$ through the first Friedmann equation. In addition, they offer an expression for the deceleration parameter, making possible the measurement of distances of different sources. Finally, these same parameters determine the age of the Universe.

3.3 Solutions to the Friedmann Equations

In this section we give exact solutions to the first Friedmann equation:

$$\dot{a}^2 = H_0^2\left(\frac{\Omega_m}{a} + \frac{\Omega_r}{a^2} + \Omega_\Lambda a^2 + \Omega_k\right) \tag{3.56}$$

While a general analytical solution to this equation does not exist, one can find some relatively simple solutions by studying special cases. For example, solutions exist for

single-component Universes, which are a good approximation for periods in cosmic time where the equation of state was approximately constant, like the periods of matter, radiation and dark energy domination. Moreover, analytical solutions exist for two-component Universes, which are interesting because they incorporate the transition era between the two equations of state.

Matter Universes: $\Omega_m + \Omega_k = 1$

We first study the case of a Universe containing no amount radiation and dark energy:

$$\dot{a}^2 = H_0^2 \left(\frac{\Omega_m}{a} + \Omega_k \right) \tag{3.57}$$

- Consider firstly a Universe where there is only matter, so that $\Omega_m = 1$ and $\Omega_k = 0$. In this case, the Friedmann equation simplifies down to the first in (3.49):

  $$\dot{a}^2 = H_0^2 \Omega_m a^{-1} \tag{3.58}$$

  Imposing the condition $a(t_0) = 1$, the solution is:

  $$a(t) = \left(\frac{t}{t_0} \right)^{2/3} \tag{3.59}$$

  This solution is historically interesting. It's called the "Einstein-de Sitter" (EdS) Universe, having been first derived by them. The age of the Universe predicted by this model is obtained solving the integral (3.55),:

  $$t_0 = \frac{2}{3} \frac{1}{H_0} \tag{3.60}$$

  Therefore, given an estimate of the Hubble constant H_0, one can easily calculate the age of the Universe. Although this solution doesn't completely describe the evolution of the Universe, it is a good approximation for the period of matter domination, as we said above.

- Consider now the case of $k = -1$, i.e. an open Universe. The solution can be expressed in parametric form:

  $$a(\theta) = \frac{1}{2} \frac{\Omega_m}{|1 - \Omega_m|} (\cosh\theta - 1) \quad t(\theta) = \frac{1}{4} \frac{\Omega_m}{|1 - \Omega_m|} (\sinh\theta - \theta) \tag{3.61}$$

  where $0 \leq \theta \leq \infty$. Note first that the solution reduces to $a(t) \propto t^{2/3}$ at early epochs $\theta \ll 1$: because of the difference in scaling between matter (a^{-3}) and curvature (a^{-2}), as long as $\Omega_m \neq 0$, there will always be an arbitrarily early time when matter dominates. More importantly, at later epochs $\theta \gg 1$ and $a \propto t$: a phase of free expansion of the Universe which will continue forever.

- Finally, consider the solution for $k = +1$, given in parametric form by:

  $$a(\theta) = \frac{1}{2} \frac{\Omega_m}{|1 - \Omega_m|} (1 - \cos\theta) \quad t(\theta) = \frac{1}{4} \frac{\Omega_m}{|1 - \Omega_m|} (\theta - \sin\theta) \tag{3.62}$$

where $0 \leq \theta \leq 2\pi$. Interestingly, in this case the Universe reaches a maximum size $a_{\max}$ at time $t_{\max}$, given by:

$$a_{\max} = \frac{\Omega_m}{|1-\Omega_m|} \quad t_{\max} = \frac{\pi}{4}\frac{\Omega_m}{|1-\Omega_m|} \tag{3.63}$$

After reaching this maximum size, the Universe recollapses in a "big crunch" at $\theta = 2\pi$, where:

$$a_{\text{crunch}} = 0 \quad t_{\text{crunch}} = \frac{\pi}{2c}\frac{\Omega_m}{|1-\Omega_m|} \tag{3.64}$$

Again, note that at early epochs we have $a \propto t^{2/3}$, for the same reason as before.

The following plot summarises the different cases:

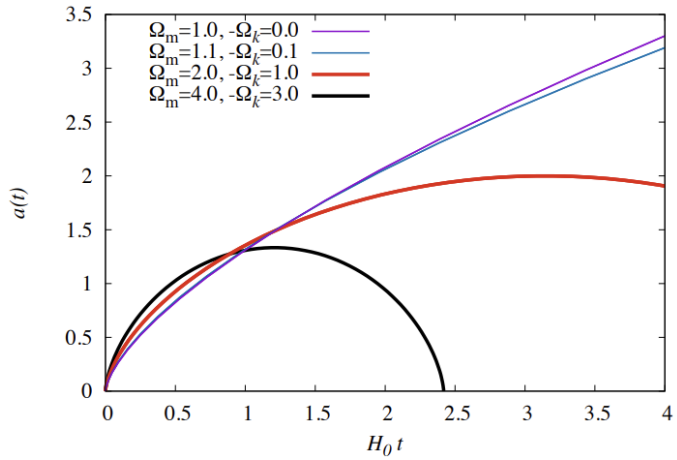

FIGURE 3.1: The three cosmologies where $\Omega_m + \Omega_k = 1$, corresponding to flat, open and closed Universes.

As was derived, a flat geometry ($\Omega_m = 1$ and $k = 0$) entails an ever expanding Universe with "escape velocity". On the other hand, an open Universe ($\Omega_m < 1$ and $k = -1$) still expands forever, whereas in a closed Universe ($\Omega_m > 1$ and $k = +1$) the matter density is large enough that it recollapses on itself.

Radiation Universes: $\Omega_r + \Omega_k = 1$

Consider now solutions for a radiation Universe void of both matter and dark energy, for which (3.39) reduces to:

$$\dot{a}^2 = H_0^2\left(\frac{\Omega_r}{a^2} + \Omega_k\right) \tag{3.65}$$

- As for the matter Universes, let's analyse first the case of zero spatial curvature $k = 0$:

$$\dot{a}^2 = H_0^2\Omega_r a^{-2} \tag{3.66}$$

Again, imposing the condition $a(t_0) = 1$, we have the following solution:

$$a(t) = \left(\frac{t}{t_0}\right)^{1/2} \tag{3.67}$$

Calculating the age of the Universe with the integral (3.55), one gets:

$$t_0 = \frac{1}{2}\frac{1}{H_0} \tag{3.68}$$

This solution is particularly interesting for the very early Universe. As said above, in fact, as long as $\Omega_r \neq 0$, there will always be an arbitrarily early time such that radiation dominates, because of its scaling as a^{-4}.

- Next, consider some amount of curvature such that $k = +1$, i.e. a closed Universe. This case is very similar to the closed Universe with matter, in the sense that it has a big crunch at some finite time in the future:

$$a(\theta) = \sqrt{\frac{1}{2}\frac{\Omega_r}{|1-\Omega_r|}}\sin\theta \quad t(\theta) = \frac{1}{2}\sqrt{\frac{1}{2}\frac{\Omega_r}{|1-\Omega_r|}}(1-\cos\theta) \tag{3.69}$$

Where, again, $0 \leq \theta \leq 2\pi$. Notice that the crunch time, and therefore the maximum, is reached at an earlier time compared to the matter Universe:

$$a_{\text{curnch}} = 0 \quad t_{\text{crunch}} = \sqrt{\frac{1}{2}\frac{\Omega_r}{|1-\Omega_r|}} \tag{3.70}$$

Basically, in both models, the Universe expands under a cycloidal evolution of the scale factor, where a maximum size is obtained at t_{max} after which the expansion is reversed. The Universe starts to compress until it reaches the big crunch at t_{crunch}.

- Finally, consider a radiation dominated open Universe, with $k = -1$. In this case, the expansion will go on forever:

$$a(\theta) = \sqrt{\frac{1}{2}\frac{\Omega_m}{|1-\Omega_r|}}\sinh\theta \quad t(\theta) = \frac{1}{2}\sqrt{\frac{1}{2}\frac{\Omega_r}{|1-\Omega_r|}}(\cosh\theta - 1) \tag{3.71}$$

where $0 \leq \theta \leq \infty$. The early Universe limit $\theta \ll 1$ correctly reproduces the zero curvature solution $a \propto t^{1/2}$, whereas the late Universe $\theta \gg 1$ sees an expansion $a \propto t$ that will continue forever.

Dark energy Universes: $\Omega_\Lambda + \Omega_k = 1$

Consider a Universe with a cosmological constant $\Lambda > 0$ and some curvature:

$$\dot{a}^2 = H_0^2\left(\Omega_\Lambda a^2 + \Omega_k\right) \tag{3.72}$$

This kind of Universe is a good approximation of late and present times, where dark energy starts to dominate all other forms of energies. The solutions to this equation are:

$$a(t) = \sqrt{\frac{3}{\Lambda}}\begin{cases}\cosh(\sqrt{\Lambda c^2/3}t) & k=+1\\ \exp(\sqrt{\Lambda c^2/3}t) & k=0\\ \sinh(\sqrt{\Lambda c^2/3}t) & k=-1\end{cases} \tag{3.73}$$

Note that the $k = +1$ solution doesn't have a singularity, whereas the scale factor vanishes at $t = -\infty$ and $t = 0$ for the $k = 0$ and $k = -1$ solutions respectively. These singularities are, however, only due to poor coordinate choices, since the three solutions (3.73) are only three different ways to slice the the same "de Sitter" space.

Early Universe: matter and radiation

As will be illustrated in the next chapter, a flat ($k = 0$) Universe containing matter and radiation is very close to resembling our own, at least at early stages. The Friedmann equation that needs to be solved is:

$$\dot{a}^2 = H_0^2 \left(\frac{\Omega_m}{a} + \frac{\Omega_r}{a^2} \right) \tag{3.74}$$

This model presents a transition from a period of radiation domination to a period of matter domination. The transition time, or scale factor a_{eq}, called "matter-radiation equality" is defined by:

$$\frac{\Omega_m}{a_{\text{eq}}} = \frac{\Omega_r}{a_{\text{eq}}^2} \implies a_{\text{eq}} = \frac{\Omega_r}{\Omega_m} \tag{3.75}$$

Rewriting (3.74) using this new definition gives:

$$\dot{a}^2 = \frac{H_0^2 \Omega_r}{a^2} \left(1 + \frac{a}{a_{\text{eq}}} \right) \tag{3.76}$$

This implies the following differential equation:

$$H_0 dt = \frac{a}{\Omega_r^{1/2}} \left(1 + \frac{a}{a_{\text{eq}}} \right)^{-1/2} da \tag{3.77}$$

Integrating it, the solution reads:

$$H_0 t = \frac{4 a_{\text{eq}}^2}{3 \Omega_r^{1/2}} \left(1 - \left(1 - \frac{a}{2 a_{\text{eq}}} \right) \left(1 + \frac{a}{a_{\text{eq}}} \right)^{1/2} \right) \tag{3.78}$$

As a sanity check, it is useful to compute the radiation and matter domination (RD and MD) limits and make sure they correspond to (3.67) and (3.59). Consider first the RD limit, obtained by $a/a_{\text{eq}} \ll 1$, under which the following scaling holds:

$$a \simeq (3 H_0 \Omega_r^{1/2} t)^{1/2} \propto t^{1/2} \tag{3.79}$$

This is indeed the right proportionality. Now, the MD limit $a/a_{\text{eq}} \gg 1$ implies:

$$a = \left(\frac{3 \Omega_r H_0 t}{4 a_{\text{eq}}^{1/2}} \right)^{2/3} \propto t^{2/3} \tag{3.80}$$

Again, this coincides with the Einstein-de Sitter solution. As a final application, one can also calculate the time of matter-radiation equality t_{eq}, which is found by setting

$a = a_{\rm eq}$ in (3.78):

$$t_{\rm eq} = \frac{4a_{\rm eq}^2}{3\Omega_r^{1/2}}\left(1 - \frac{1}{\sqrt{2}}\right) \tag{3.81}$$

Late Universe: matter and dark energy

Consider now a flat Universe containing matter and a positive cosmological constant, such that $\Omega_m + \Omega_\Lambda = 1$. This model describes rather well our own Universe at present and future times. The Friedmann equation to solve is, in this case:

$$\dot{a}^2 = H_0^2\left(\frac{\Omega_m}{a} + \Omega_\Lambda a^2\right) \tag{3.82}$$

The time of interest is now the moment of "matter-dark energy equality" $a_{\Lambda\rm m}$, defined by:

$$\frac{\Omega_m}{a_{\Lambda\rm m}} = \Omega_\Lambda a_{\Lambda\rm m}^2 \implies a_{\Lambda\rm m} = \left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{1/3} \tag{3.83}$$

Rewriting the Friedmann equation in the following way:

$$\dot{a} = H_0 a \Omega_\Lambda^{1/2}\left(\frac{a_{\Lambda\rm m}^3}{a^3} + 1\right)^{1/2} \tag{3.84}$$

yields the differential equation:

$$H_0 dt = \frac{a}{\Omega_\Lambda^{1/2}}\left(\frac{a_{\Lambda\rm m}^3}{a^3} + 1\right)^{-1/2} da \tag{3.85}$$

The solution is the following:

$$H_0 t = \frac{2}{3\Omega_\Lambda^{1/2}} \log\left[\left(\frac{a}{a_{\Lambda\rm m}}\right)^{3/2} + \left(\frac{a^3}{a_{\Lambda\rm m}^3} + 1\right)^{1/2}\right] \tag{3.86}$$

It is useful once again to check whether this solution gives the right scalings at early (3.59) and late (3.73) times. First, the limit $a/a_{\Lambda\rm m} \ll 1$ in (3.86) yields:

$$a \simeq \left(\frac{3H_0 t a_{\Lambda\rm m}^{3/2}\Omega_\Lambda^{1/2}}{2} t\right)^{2/3} \tag{3.87}$$

which is again the right proportionality for a matter dominated Universe. The late limit $a/a_{\Lambda\rm m} \gg 1$ instead results in:

$$a \simeq \frac{a_{\Lambda\rm m}}{2^{2/3}} \exp\left(\frac{3H_0\Omega_\Lambda^{1/2} t}{2}\right) \tag{3.88}$$

which corresponds with the result of a flat de Sitter Universe. As before, the time of matter-dark energy equality $t_{\Lambda\rm m}$ can be found by setting $a = a_{\Lambda\rm m}$ in (3.86):

$$t_{\Lambda\rm m} = \frac{2}{3H_0\Omega_\Lambda^{1/2}} \log(1 + \sqrt{2}) \tag{3.89}$$

Chapter 4

Observational Cosmology

To finally complete the picture of the Universe outlined in the chapters above, it remains to explore the observational data and figure out how it holds up when compared to the mathematical description.
The main theoretical results are the following. The energy budget is subdivided into three components: matter (baryonic and dark), radiation and dark energy. Most importantly, at large enough scales (say $\gtrsim 100$ Mpc) the Universe appears isotropic and homogeneous, a fact that helped tremendously to lay out the formalism. The metric is then very simple and assumes the form:

$$ds^2 = -c^2dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2d\Omega^2\right] \tag{4.1}$$

In turn the Einstein equations can be grouped into only one differential equation, the (a-dimensional) Friedmann equation, for the scale factor $a(t)$, which governs the dynamics of the Universe:

$$H^2 = H_0^2\left[\Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_k a^{-2} + \Omega_\Lambda\right] \tag{4.2}$$

Thanks to modern observational techniques, the adimensional density parameters are now known to a very high degree of precision. The sections that follow contain a summary of the relevant data that was extracted from observations.

The density parameters

Starting with radiation, in particular photons, their density is easily recovered from the CMB, the backround radiation that has a recorded black-body temperature of:

$$\bar{T}_0 = 2.73\ \mathrm{K} \tag{4.3}$$

Statistical mechanics provides a relation between number/energy density of a system of photons in thermodynamical equilibrium and its temperature:

$$\begin{aligned} n_{\gamma,0} &= \frac{2\zeta(3)}{\pi^2}\left(\frac{\bar{T}_0^2}{\hbar c}\right)^3 \simeq 410\ \text{photons cm}^{-3} \\ \rho_{\gamma,0} &= \frac{\pi^2}{15}\left(\frac{\bar{T}_0^2}{\hbar c}\right)^4 \simeq 4.6\times 10^{-34}\ \text{g cm}^{-3} \end{aligned} \tag{4.4}$$

(here $\zeta(s)$ is the Riemann zeta function, with $\zeta(3) \simeq 1.2$) where the factor of c was reinstated for the time being. Converting this energy density in units of critical density

ρ_0, the adimensional photon density parameter is measured to be:

$$\Omega_\gamma = \frac{\rho_{\gamma,0}}{\rho_0} \simeq 5.4 \times 10^{-5} \tag{4.5}$$

On the other hand, as far as neutrinos are concerned, as long as they are relativistic their energy density can be show to be roughly 68% that of photons. Extrapolating this assumption to the present time yields:

$$\Omega_\nu \simeq 3.6 \times 10^{-5} \tag{4.6}$$

In total, therefore, the radiation energy density is:

$$\Omega_r = \Omega_\gamma + \Omega_\nu \simeq 9 \times 10^{-5} \tag{4.7}$$

As is currently well known, however, neutrinos are massive, despite their mass being substantially small. As a consequence, they become non-relativistic at late times, which increases their energy density contribution. Current observations constrain its range to be $0.0012 < \Omega_\nu < 0.003$.

Next we analyse the matter sector, which is composed of a baryonic and a dark component. The baryonic energy density is known because of the theory of Big Bang Nucleosynthesis, which predicts the abundances of light elements that were produced during the hot Big Bang. In addition, the Ω_b determines certain features of the CMB. Coupled together, these two observations yield:

$$\Omega_b \simeq 0.05 \tag{4.8}$$

Most of the matter in the Universe, however, is in the form of dark matter, whose effects are mainly observed in the evolution and formation of large-scale structures (galaxies and clusters of galaxies) and in the CMB. The inferred energy density is:

$$\Omega_c \simeq 0.27 \tag{4.9}$$

where the subscript c indicates that we are assuming a cold form of dark matter. The total amount of matter in the Universe is thus:

$$\Omega_m = \Omega_b + \Omega_c \simeq 0.32 \tag{4.10}$$

Next, we note that the CMB actually gives precious information about the spatial curvature of the Universe. In fact, curvature changes the angular diameter distance to the surface of last-scattering, affecting the angle at which the scale is observed. These types of measurements give an upper bound to the energy density in the form of curvature:

$$|\Omega_k| < 0.005 \tag{4.11}$$

This number, therefore, suggests that the Universe is actually flat $k \simeq 0$. In other words, the spacetime metric is of the form:

$$ds^2 = -dt^2 + a^2(t)\delta_{ij}dx^i dx^j \tag{4.12}$$

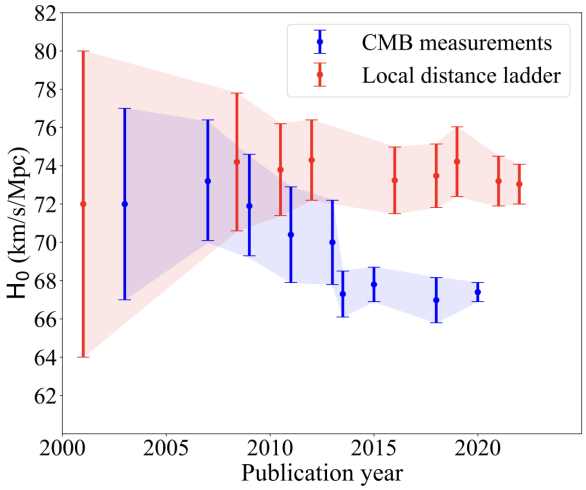

FIGURE 4.1: The Hubble tension

Finally, the presence of a dark energy component, which makes the Universe accelerate, was inferred from the brightness of distant supernovae explosions. These appeared fainter than expected in a Universe only made up of matter and radiation. Assuming a flat Universe, the data could only be fit with the following dark energy density parameter:

$$\Omega_\Lambda \simeq 0.68 \tag{4.13}$$

A further point of interests regards the H_0 parameter, the Hubble constant. It is a fundamental value which, among other things, is instrumental in determining the age of the Universe. For this very reason, it is important to determine it accurately. Since measurements of this parameter used to come with very large uncertainties, it is conventional to define it as:

$$H_0 \equiv h\, 100 \text{ km s}^{-1} \text{ Mpc}^{-1} \tag{4.14}$$

However, the accuracy of H_0 measurements has improved drastically in the last decade, and it turns out that it gives rise to a fundamental statistical discrepancy. This constant can be measured directly from the cosmological model using early Universe data extracted from the CMB. The state of the art for these observations give:

$$h = 0.674 \pm 0.005 \tag{4.15}$$

On the other hand, the same constant can be measured with local Universe data, based on the distance-redshift relation. These methods are usually undertaken by building a "local distance ladder", which involves supernovae measurements; these methods won't be explained here. Interestingly, this H_0 measurement is in stark contrast with (4.15), since it gives:

$$h = 0.73 \pm 0.010 \tag{4.16}$$

The discrepancy between these two values has been dubbed the "Hubble tension", and it is thought to be the most serious challenge to the ΛCDM model. In fact, as can be seen in Figure 4.1, the measurements have been made more and more precise with the years, ruling out possible systematic uncertainties as being the solution. For

this reason, at least part of the community thinks that the Hubble tension may be a gateway to new physics.

Once the density parameters are specified, the evolution of the Universe through its phases can be easily traced. Initial conditions are set at 100 GeV, where radiation is the dominant contribution to the energy content. After a relatively brief span of time, the matter energy density starts to become non-negligible, until it starts to dominate at the "moment" of matter-radiation equality, which we have defined previously and can now calculate:

$$\begin{aligned} a_{\text{eq}} &= \frac{\Omega_r}{\Omega_m} \simeq 2.9 \times 10^{-4} \\ z_{\text{eq}} &\simeq 3400 \end{aligned} \tag{4.17}$$

Matter dominates the Universe for a long time, until close to now at matter-dark energy equality, when the energy provided by cosmological constant term gains the upper hand:

$$\begin{aligned} a_{\text{m}\Lambda} &= \left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{1/3} \simeq 0.77 \\ z_{\text{m}\Lambda} &\simeq 0.3 \end{aligned} \tag{4.18}$$

Unless the nature of dark energy is wildly different than currently thought, the vacuum energy will dominate forever. Notice that there is no period where the energy from the spatial curvature is the leading term, due to the very negligible part of the energy budget made up by Ω_k.

Once the density parameters Ω are known, it is not only possible to discern the complex dynamics of the Universe, but also to delve even deeper into the frameworks that we have introduced along the way, such as the deceleration parameter, the age of the Universe, the concept of horizons and the significance of distances in cosmology.

Acceleration or deceleration?

The deceleration parameter is defined as:

$$q \equiv -\frac{\ddot{a}}{aH^2} \tag{4.19}$$

Since the values at the denominator are strictly positive, if $q > 0$ the Universe is decelerating and viceversa, if $q < 0$. Using the second Friedmann equation and the definition of the critical density, one gets the cleaner expression:

$$q(\{\Omega\}) = q(\{\Omega_r, \Omega_m, \Omega_\Lambda\}) = \frac{1}{2}\sum_i \Omega_i(1+3w_i) = \frac{1}{2}(\Omega_m + 2\Omega_r - 2\Omega_\Lambda) \tag{4.20}$$

Evaluating this parameter at present time:

$$q_0 = q(\{2.9 \times 10^{-4}, 0.32, 0.7\}) \simeq \frac{1}{2}(0.32 - 2 \times 0.7) < 0 \tag{4.21}$$

In other words, at present time the Universe is accelerating. Since dark energy will dominate forever, at least in so far as our current scientific understanding is concerned, this acceleration will continue forever.
On the other hand, during the radiation (RD) and matter (MD) dominated periods, things were different. Since the sign of q is the only important parameter, one can draw conclusions without the need to calculate the Ωs for these two periods. In fact, these periods can be assigned the following set of parameters: RD $\simeq \{1,0,0\}$ and MD $\simeq \{0,1,0\}$, since Ω_r and Ω_m are respectively the dominating terms. Then:

$$\begin{aligned} \mathrm{sgn}(q_{\mathrm{RD}}) &= \mathrm{sgn}(q(\{1,0,0\})) = +1 \\ \mathrm{sgn}(q_{\mathrm{MD}}) &= \mathrm{sgn}(q(\{0,1,0\})) = +1 \end{aligned} \tag{4.22}$$

Therefore, when the dominating energy was either that of radiation or matter, the Universe was decelerating.

The age of the Universe

The concept of acceleration and deceleration during cosmic expansion also affects the deduced age of the Universe. The function $t(a)$, as introduced above, reads:

$$t(a,\{\Omega\}) = \frac{1}{H_0}\int_0^a \frac{da}{[\Omega_r a^{-2} + \Omega_m a^{-1} + \Omega_\Lambda a^2 + \Omega_k]^{1/2}} \tag{4.23}$$

Now, the age of the Universe is defined as $t_0 \equiv t(1)$, which results in:

$$t_0 = t(\{2.9\times10^{-4}, 0.32, 0.7\}) \simeq 13.8\times10^9 \text{ years} \tag{4.24}$$

Finally, it is also interesting to calculate the times of matter-radiation and matter-dark energy equalities by setting $t_{\mathrm{eq}} = t(a_{\mathrm{eq}})$ and $t_{\mathrm{m}\Lambda} = t(a_{\mathrm{m}\Lambda})$. The integral gives:

$$\begin{aligned} t_{\mathrm{eq}} &\simeq 50,000 \text{ years} \\ t_{\mathrm{m}\Lambda} &\simeq 10.2\times10^9 \text{ years} \end{aligned} \tag{4.25}$$

The Hubble horizon

The notion of a horizon is fundamental in Cosmology. We have introduced the particle horizon $d_{\mathrm{h}}(t)$, which is the maximum distance light could have traveled from the beginning of the Universe up until time t, and the event horizon $d_{\mathrm{e}}(t)$, the maximum distance light could travel from time t to the infinite future. These two are defined in the following way:

$$\begin{aligned} d_{\mathrm{h,\,comoving}}(\eta) &= \chi = \eta = \int_0^t \frac{dt'}{a(t')} \\ d_{\mathrm{e,\,comoving}}(\eta) &= \int_\eta^{\eta_f} d\eta = \eta_f - \eta = \int_t^\infty \frac{dt'}{a(t')} \end{aligned} \tag{4.26}$$

In a diagram where the time t is replaced with the conformal time η, these two horizons are easy to draw, as they are straight lines. Finally, the notion of Hubble horizon was introduced as the distance at which objects recede at the speed of light.

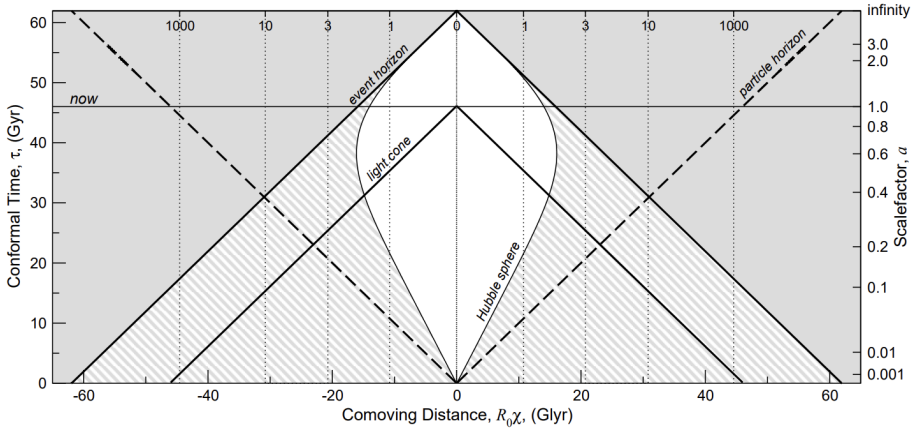

FIGURE 4.2: The conformal diagram showing the horizons.

The comoving Hubble sphere is:

$$d_{\mathrm{HR}}(t) = \frac{1}{a(t)H(t)} \tag{4.27}$$

Knowing the form of $a(t)$, we can now plot this horizon too. First, it is easier to convert $d_{\mathrm{HR}}(t)$ to conformal time, using the following relations:

$$a(t) \propto \begin{cases} t^{1/2} & \mathrm{RD} \\ t^{2/3} & \mathrm{MD} \\ e^{H_0\sqrt{\Omega_\Lambda}t} & \Lambda\mathrm{D} \end{cases} \tag{4.28}$$

Using the definition $\eta = \int_0^t dt/a$, one finds:

$$a(\eta) \propto \begin{cases} \eta & \mathrm{RD} \\ \eta^2 & \mathrm{MD} \\ -\frac{1}{\eta} & \Lambda\mathrm{D} \end{cases} \tag{4.29}$$

The comoving Hubble horizon in conformal time is thus:

$$d_{\mathrm{HR}}(\tau) = \frac{a(\eta)}{a'(\eta)} \propto \begin{cases} \eta & \mathrm{RD} \\ \eta & \mathrm{MD} \\ -\eta & \Lambda\mathrm{D} \end{cases} \tag{4.30}$$

This is the mathematical perspective of the arguments put forward above. The Hubble sphere gets bigger if the Universe is expanding in a decelerating fashion, such as during radiation and matter dominated periods. This is because the distance at which objects recede at the speed of light grows, which is precisely the definition of an increasing Hubble horizon. On the other hand, an accelerated expansion, such as that caused by vacuum energy, shrinks the Hubble sphere, such that objects recede faster progressively close to us. The Hubble sphere is depicted in Figure 4.2, along with the particle and event horizons. Figure 4.2 also encapsulates neatly this discussion. There is a way of being outside the Hubble sphere and inside the past light cone, like at the bottom of the plot, when the Universe is still decelerating. In

this way an observer who founds himself outside the Hubble horizon can communicate with us in the future. Clearly, this is not possible when dark energy starts to dominate, as the top of the figure displays. The light cone closes up following the Hubble sphere, so that no observer outside it will be able to send us any information, forever.

Distances and the surface brightness

In cosmology, measuring the distance of an object can be quite a difficult task. One can make use of its apparent luminosity (the energy flux detected on Earth), and then convert it to a distance using its intrinsic luminosity, provided one knows it. This definition is the basis of the luminosity distance $d_L(z)$, which was found to be given by the following series:

$$d_L(z) = \frac{c}{H_0}\left[z + \frac{1}{2}(1 - q_0)z^2 + \dots\right] \tag{4.31}$$

On the other hand, the object's angular size can be used to infer its distance. This line of reasoning brought us instead to define the angular diameter distance $d_A(z)$:

$$d_A(z) = \frac{c}{H_0}\left[z - \frac{1}{2}(1 + q_0)z^2 + \dots\right] \tag{4.32}$$

Using the values of the density parameters Ω, the following plot displays the two series:

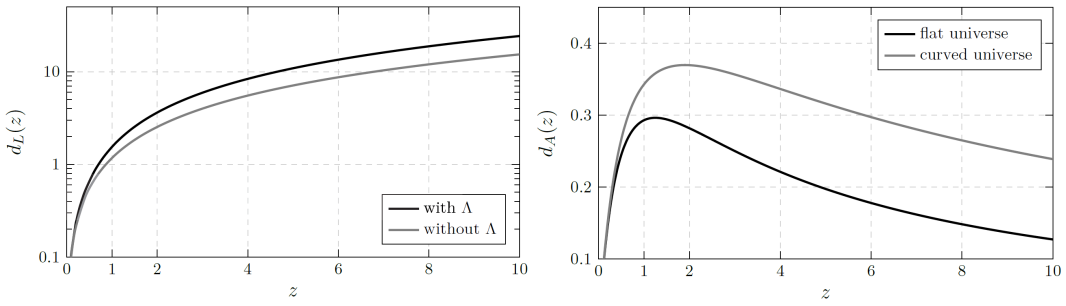

FIGURE 4.3: Distances in cosmology

Notice that the angular diameter distance exhibits apparently paradoxical behavior, that is there exists a turning point such that, beyond it, objects appear larger the more distant they are. In other words, if one looks at galaxies at increasing redshifts, those that reside at greater redshifts appear bigger in the sky (remember that $\delta\theta \sim 1/d$). This turnover point occurs because the Universe is expanding. Objects that are now very distant actually were closer to us in the past. The light that we receive from these now-distant objects was emitted by them a long time ago, when they were nearer, and therefore spanned a larger angle in the sky.
Luckily, there is a way of distinguishing between the distance two objects that appear to be roughly the same dimension in the sky, but are at different redshifts. It has to do with a quantity called the surface brightness $\mathcal{B}$, which is defined (for extended

objects) as the ratio of an object's apparent luminosity and its area:

$$\mathscr{B} \equiv \frac{\ell}{A} \tag{4.33}$$

In Euclidian optics, both quantities scale as d^{-2}, d being the distance to the object. For this reason one usually says that surface brightness is conserved. This does not happen in cosmology, where in fact:

$$\mathscr{B} \propto \frac{d_L^{-2}}{d_A^{-2}} \propto (1+z)^{-4} \tag{4.34}$$

This effect is called "cosmological dimming". This solves the problem of the turnaround point of the angular diameter distance, since in reality objects actually appear fainter as their redshifts increase. This phenomenon is a big part of the reason why it is hard to see objects that are really young, that is at $z \gg 1$.

Chapter 5

Inflation

Large-scale homogeneity and flatness are the most fundamental and important concepts in cosmology. The implications of these simple concepts are however problematic.

5.1 The horizon problem

It is clear from the Cosmic Microwave Background that the observable components of the very early Universe, photons and baryons, constituted a system that was very uniform. The CMB is a perfect black body with temperature $\bar{T}_0 \sim 2.73$ K. How is it so? How was the Universe so uniform at such an early stage? How can two patches chosen at random in the sky possess the same temperature? The obvious explanation that comes to mind is thermalization. If we think of a gas of particles in a box with random initial conditions, the system will eventually relax to a state of maximum entropy with a shared temperature, that is the particles will find themselves in thermodynamical equilibrium. Applying the same idea to the Universe, however, fails miserably. Different parts of the CMB were so far apart at the time of recombination (when the CMB was formed) that they were not even in causal contact with each other. In other words, the particle horizon (the maximum distance light could have traveled from $t = 0$) at recombination was smaller than the distances of some patches with the same temperature in the sky. The "horizon problem" is then the following: how could causally disconnected patches have evolved to have the same exact temperature?
The problem can be quantified rather precisely. Recall the definition of the (comoving)particle horizon as the logarithmic integral of the (comoving)Hubble sphere:

$$d_h(\eta) = \int_{\log a_i}^{\log a} (aH)^{-1} d\log a \tag{5.1}$$

To solve the integral, we first derive the expression for the Hubble horizon in a flat Universe with only matter and radiation. The Friedmann equation gives us the form of the Hubble parameter:

$$H^2 = H_0^2[\Omega_m a^{-3} + \Omega_r a^{-4}] \tag{5.2}$$

where the time of equality is $a_{\text{eq}} = \Omega_r/\Omega_m \simeq 3400^{-1}$. The Hubble sphere is then given by:

$$(aH)^{-1} = \frac{1}{\sqrt{\Omega_m}} H_0^{-1} \frac{a}{\sqrt{a + a_{\text{eq}}}} \tag{5.3}$$

Therefore, the particle horizon is the following:

$$d_h(a) = \int_0^a \frac{d\log a}{aH} = \frac{2}{\sqrt{\Omega_m}} H_0^{-1}(\sqrt{a + a_{\rm eq}} - \sqrt{a_{\rm eq}}) \tag{5.4}$$

Evaluating the horizon today ($a_0 = 1$) and at recombination ($z_\star \simeq 1100$, $a_\star \simeq 1100^{-1}$):

$$\begin{aligned} d_{h,0} &= \eta_0 \simeq \frac{2}{\sqrt{\Omega_m}} H_0^{-1} \simeq 20,285 \text{ Mpc} \\ d_{h,\star} &= \eta_\star = \frac{2}{\sqrt{\Omega_m}} H_0^{-1}[\sqrt{1100^{-1} + 3400^{-1}} - \sqrt{3400^{-1}}] \simeq 265 \text{ Mpc} \end{aligned} \tag{5.5}$$

Using the comoving distance to the surface of last scattering (recombination) $\chi_\star = \eta_\star - \eta_0 \simeq 15.1$ Gpc, we can estimate easily the angle subtended by the particle horizon at recombination:

$$\theta_h = \frac{2d_{h,\star}}{\chi_\star} \simeq 0.036 \text{ rad} \simeq 2.0^\circ \tag{5.6}$$

This neat results shows the extent of the horizon problem. Patches in the CMB that are more than 2° apart could not have interacted before recombination. Dividing the total square degrees on the sphere by 2° squared, we find that there are roughly 40000 causally disconnected patches in the CMB.
The following pictures depict the problem in a straightforward way:

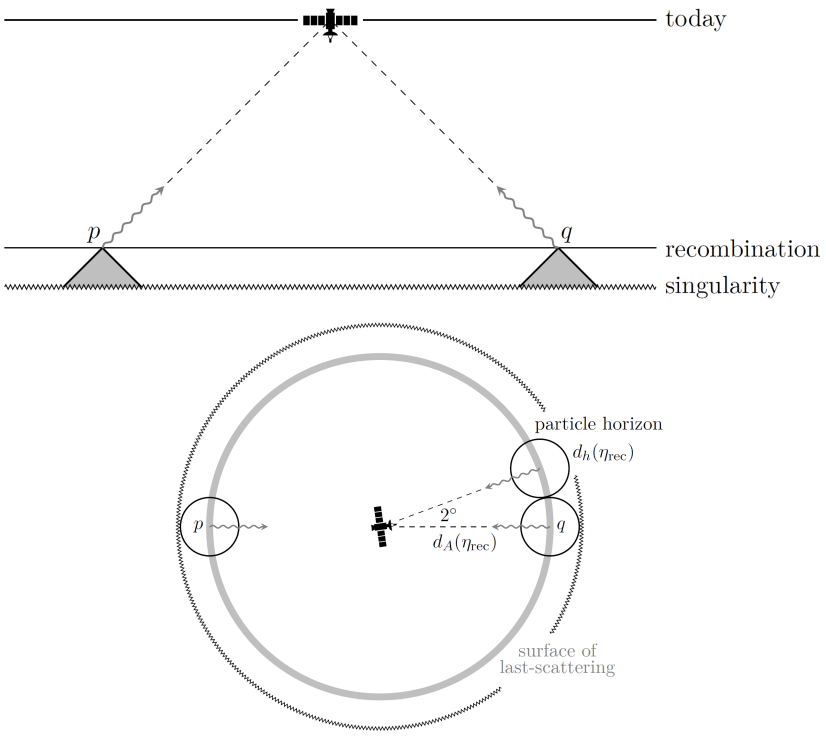

FIGURE 5.1: *Top*: Conformal diagram showing the light cones. *Bottom*: Pictorial representation of the particle horizons.

The picture on top shows two points in the sky, p and q, separated by 180 degrees. Since the particle horizon today η_0 (half of the base of the big triangle) is

orders of magnitudes bigger than at recombination $\eta_\star$ (half of the base of the shaded triangle), the light cones of p and q do not overlap. They were never in causal contact before $\eta_\star$. Notice that we could have carried out the same argument if p and q were separated by only a little more than 2 degrees. Furthermore, the situation is even worse than depicted since $\eta_\star \simeq 10^{-2}\eta_0$, meaning that the shaded triangles should be even smaller. The picture on the bottom shows instead that CMB scales span much larger distances than the particle horizon at the surface of last scattering.

The horizon problem is exacerbated by the fact that the CMB actually shows small perturbations in its temperature, of the order of $\delta T/\bar{T}_0 \sim 10^{-5}$. Notwithstanding our origin of these perturbations, the fundamental problem resides in their correlations spanning acausal distances. To understand the problem, let's look at the figure below:

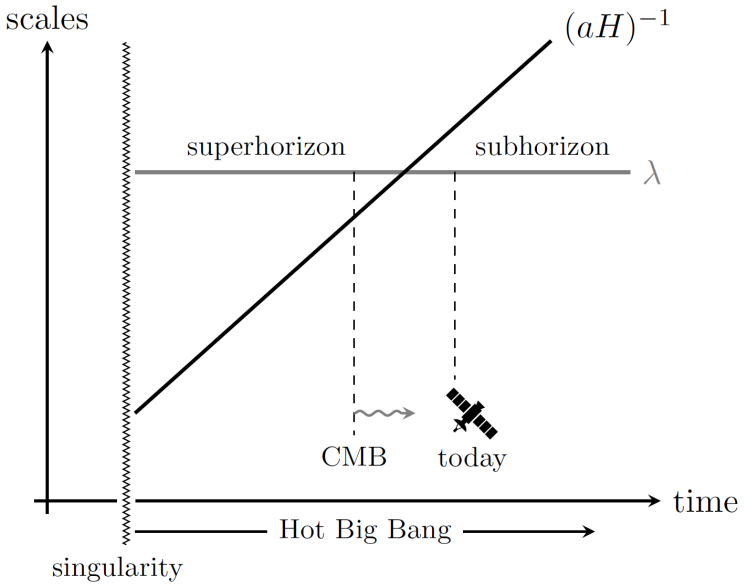

FIGURE 5.2: Comoving scales exiting the horizon at early times

The plot shows the evolution of a scale λ and the Hubble horizon $(aH)^{-1}$ in comoving coordinates. The scale, being comoving, stays the same, whereas the Hubble sphere grows, as should be clear from the derived expression (5.3). This implies that any scale that might reside inside the horizon today (subhorizon) will have found itself outside of it (superhorizon) at some early time. Then, the problem is the following. Although the horizon at recombination is about 265 Mpc, observations reveal that CMB fluctuations are correlated on scales larger that this value, even though they should not have been able to communicate in any way.
Therefore, not only is the CMB thermalised across acausal distances, but on these scales it also displays correlations between fluctuations.

Beware of the confusion

There is a subtle point that is worth discussing before moving forward. In fact, a fundamental concept in cosmology is that of (physical) scales entering and exiting the horizon, since only subhorizon modes are observable and are causally connected. In the discussion just above, this horizon was implied to be the Hubble sphere. But wait! Doesn't the particle horizon decide whether two points in spacetime are causally related or not? The resolution of this confusion lies in noticing that the particle and Hubble horizons are actually proportional to each other during both radiation and matter domination periods. For this reason, textbooks use the two concepts interchangeably.
To verify this claim, take the scale factor to be $a(t) \propto t^n$, where $n = 1/2, 2/3$ for radiation and matter domination respectively. The particle horizon then is:

$$d_h(t) = \int_0^t \frac{dt'}{a(t')} \propto \int_0^t dt' t'^{(-n)} = \frac{1}{1-n} t^{1-n} \tag{5.7}$$

On the other hand, the Hubble sphere grows like:

$$(aH)^{-1}(t) = \dot{a}^{-1}(t) \propto t^{1-n} \tag{5.8}$$

In this sense we see that up to numerical factors the two horizons are the same, that is $d_h(t) \propto (aH)^{-1}(t)$.
Since every (comoving)scale λ can be identified with its (comoving)wavenumber, $k = 2\pi/\lambda$, we have the following rule:

$$\begin{aligned} k \ll aH &\longleftrightarrow \text{The scale } \lambda \text{ is SUPERHORIZON} \\ k \gg aH &\longleftrightarrow \text{The scale } \lambda \text{ is SUBHORIZON} \end{aligned} \tag{5.9}$$

We say that a mode "enters the horizon" as it goes from $k \ll aH$ to $k > aH$, whereas it "leaves the horizon" if it goes from $k \gg aH$ to $k < aH$. Figuring out whether a certain scale is inside or outside the horizon is especially important since only inside the horizon can it become observable. Furthermore, this framework is critical in cosmological perturbation theory (the study of the evolution of the small perturbations around uniformity), because a given perturbation on a scale with wavenumber k is frozen in time if $k < aH$, i.e. if it lies outside the horizon, as causality prevents it from growing.

It is worth stressing that while textbooks replace the concept of particle horizon with that of the Hubble sphere, the latter is independently related to causality. In fact, the Hubble radius is the approximate distance over which light can travel in the course of one expansion time, i.e. the time it takes for the scale factor to increase by a factor of e. In other words, the Hubble horizon provides a yardstick to measure whether two observers can interact within one expansion time.
There is, however, a fundamental difference between the two concepts. If two observers are separated by a distance greater than the Hubble sphere, they cannot communicate with each other now (that is in the course of one expansion time, as was explained above), but it is possible that they have communicated in the past. This could happen, for example, if the Hubble sphere had a period of rapid decrease, as is the case with inflation (see later). Instead, if the two observers are at some time

t separated by a distance greater than the particle horizon, they could have never communicated before time t. This should be clear, since the particle horizon (the conformal time) traces the observer's light cone.

5.2 The flatness problem

Precise cosmic microwave background measurements have provided an upper bound on the value of today's curvature parameter $|\Omega_k| < 5 \times 10^{-3}$. This value, however, requires extremely fine-tuned initial conditions. To see why, consider the Friedmann equation for a Universe with unknown curvature:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \equiv \frac{8\pi G}{3}(\rho + \rho_k) \tag{5.10}$$

Using the definition of critical density $\rho_c(t) = 3H^2/8\pi G$, we get:

$$1 = \Omega(t) + \Omega_k(t) \tag{5.11}$$

The adimensional curvature parameter thus varies in time like:

$$\Omega_k(t) = -\frac{k}{a^2H^2} = \frac{H_0^2}{(aH)^2}\Omega_k \tag{5.12}$$

The "flatness problem" arises because, as one can see, $\Omega_k(t) = 0$ is a fixed point. In fact, using our solution for the Hubble radius in (5.3) one finds:

$$\Omega_k(t) = \frac{\Omega_k}{\Omega_m}\frac{a^2}{a + a_{\text{eq}}} \tag{5.13}$$

This expression implies that the curvature of the Universe decreases when going back in time. If $\Omega_k \simeq 0$ today, in the past it should have been extremely close to zero, but not exactly zero (otherwise it would still be zero today), which is a coincidence that appears rather unlikely. To quantify this, it's just sufficient to evaluate (5.13) at some early times. Choosing these to be $z_{\text{EW}} \simeq 10^{15}$, the time of the electroweak phase transition, and $z_{\text{BBN}} \simeq 10^8$, the time of Big Bang Nucleosynthesis, we have:

$$\begin{aligned} |\Omega_k(t_{\text{EW}})| &\lesssim 10^{-29} \\ |\Omega_k(t_{\text{BBN}})| &\lesssim 10^{-16} \end{aligned} \tag{5.14}$$

These extremely small values seem artificial enough that the flatness problem is also called the "fine-tuning" problem.

5.3 The solution: a period of accelerated expansion

The causality issues that arose in the standard Friedmann Universe point towards their own solution. The horizon problem, for instance, implies that the particle horizon should be orders of magnitude bigger than what we calculate it to be. Can't we then just think of a mechanism that would increase this quantity?
An obvious candidate that could generate such a mechanism is the Hubble sphere,

since it is related to both the particle horizon and the causality structure. To build on this idea, recall the scaling of the Hubble horizon in figure 5.2. During both matter and radiation domination, the Hubble horizon's growth leads to subhorizon scales today necessarily being superhorizon at some point in the distant past, giving rise to the causality problems we discussed. In other words, this period in the Universe sees the Hubble sphere growing faster than physical scales:

$$\frac{d}{dt}\left(\frac{\lambda}{(aH)^{-1}}\right) < 0 \tag{5.15}$$

It it thus possible to think of an era, sometime before radiation domination, when this condition was reversed, such that the Hubble sphere was shrinking. This period in the history of the Universe must surely have occurred before the era of Big Bang Nucleosynthesis, since the latter is the earliest era from which we have observational data agreeing rather well with our theory, which in turn presupposes the radiation equation of state. However, the thermal history before nucleosynthesis is unknown.

The shrinking Hubble sphere solves the isotropy of the CMB rather straightforwardly. Scales that were outside of the horizon long ago were actually inside of it at the time of this primordial epoch. If this happens, the photons we see as a microwave background emitted from the last scattering surface by causally disconnected patches, actually have the same temperature because they were in causal contact (were inside the horizon) during this primordial era. Even the problem of correlated superhorizon correlations is now solved, since all scales could communicate. It is possible to visualize such a period in the Universe by extending figure 5.2 to the left in a symmetric fashion:

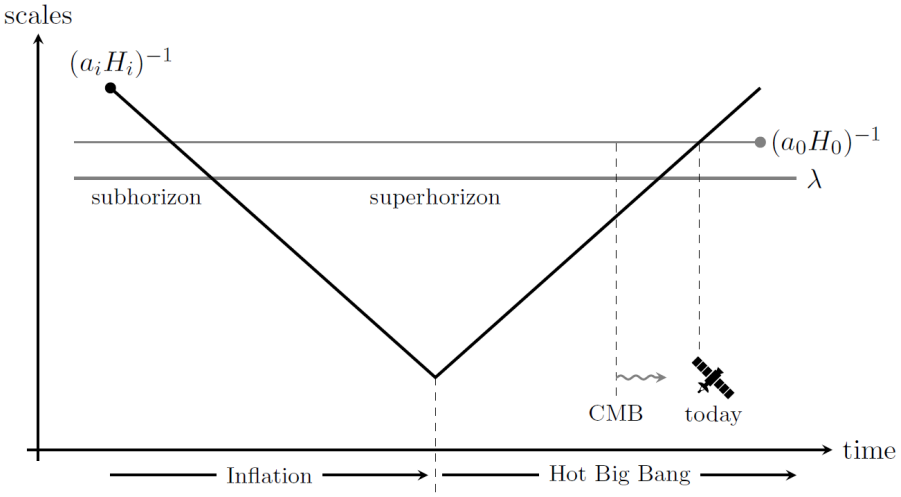

FIGURE 5.3: Scales re-entering the horizon during the period of a shrinking Hubble sphere

Notice that condition (5.15) is reversed in this new epoch:

$$\frac{d}{dt}\left(\frac{\lambda}{(aH)^{-1}}\right) > 0 \tag{5.16}$$

implying (since λ is comoving):

$$\ddot{a} > 0 \tag{5.17}$$

This period is called "inflation". It is a period of accelerated expansion that precedes the Big Bang and during which subhorizon scales were pushed outside the Hubble horizon. As we have changed the evolution of the Hubble horizon, now the particle horizon is way bigger than calculated in the Friedmann Universe, which was precisely our goal in the beginning.
An important application of the inflationary condition (5.17) is immediately found by inspection of the second Friedmann equation:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \tag{5.18}$$

Imposing $\ddot{a} > 0$ gives:

$$\rho + 3P < 0 \iff w_{\text{inflation}} < -\frac{1}{3} \tag{5.19}$$

The object that drives inflation needs to respect this condition on the equation of state. We already know from the study of dark energy that, in order to obtain an accelerated expansion, we need the pressure to be negative. It appears therefore that inflation was driven by a similar type of energy. It cannot be ordinary matter or radiation, since both have a positive equation of state, but the cosmological constant itself is also prohibited as it would give rise to a never ending inflationary period, while what is required is for inflation to end and transition to the radiation dominated phase. While the object driving inflation is still unknown, we will shortly provide the most commonly accepted framework.

Inflation and the horizon problem

From a conceptual standpoint, the horizon problem is solved. Superhorizon scales at the time of photon decoupling were subhorizon some time during the epoch of inflation. In this section we would like to offer a more detailed and quantitative description of the solution to the horizon problem.
Consider again the expression for the particle horizon:

$$d_h(\eta) = \int_{a_i}^{a} (aH)^{-1} d\log a \tag{5.20}$$

In the standard Friedmann Universe, as the Hubble horizon $(aH)^{-1}$ grows, the largest contribution to the integral comes from late times and we have $d_h \sim (aH)^{-1}$ as was previously argued. However, adding a period of a shrinking Hubble sphere changes things dramatically, because the contribution close to early times in the integral dominates, causing the particle horizon to be much bigger $d_h \simeq (a_i H_i)^{-1} \gg (aH)^{-1}$.
To illustrate this concept, consider a flat Universe dominated by a fluid with a constant equation of state w, such that the Friedmann equation is:

$$H^2 = \frac{8\pi G}{3} a^{-3(1+w)} = a^{-3(1+w)} H_0^2 \tag{5.21}$$

where we have used the density scaling of the said fluid $\rho(t) = \rho(t_0)a^{-3(1+w)}$ we derived in (4.23) and $\Omega = 1$, since it is the only fluid in the Universe. Multiplying (5.21) by a^2, we see that the Hubble horizon then evolves like:

$$(aH)^{-1} = H_0^{-1} a^{\frac{1}{2}(1+3w)} \tag{5.22}$$

Plugging this into (5.20) yields the form of the particle horizon:

$$d_h(\eta) = \eta - \eta_i = \frac{2H_0^{-1}}{(1+3w)} \left[a^{\frac{1}{2}(1+3w)} - a_i^{\frac{1}{2}(1+3w)} \right] \tag{5.23}$$

It is crucial to analyze the initial conformal time η_i:

$$\eta_i = \frac{2H_0^{-1}}{(1+3w)} a_i^{\frac{1}{2}(1+3w)} \tag{5.24}$$

In the standard Friedmann cosmology, we have set the initial singularity at $a_i = 0$ which implies $\eta_i = 0$ because $1 + 3w > 0$. Adding inflation changes this fact entirely. We see that the initial singularity $a_i = 0$, for $1 + 3w < 0$, is pushed to (negative) infinite conformal time $\eta_i \to -\infty$, causing the particle horizon to be infinite. Clearly, this is an extrapolation due to having extended the integral (5.20) to $a_i = 0$, which is questionable at best. Today, we still don't know exactly when inflation happened and how long it lasted.
However, pushing the singularity to infinite conformal time solves the problem neatly as the following picture shows:

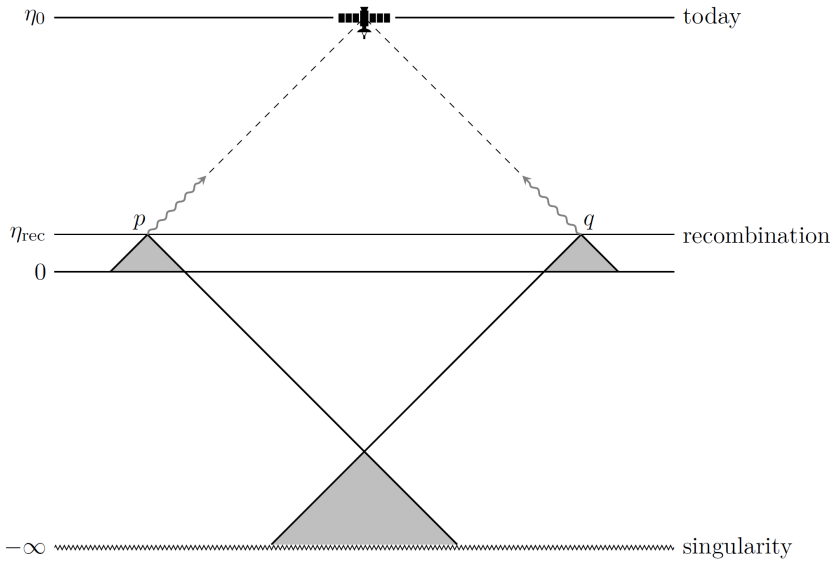

FIGURE 5.4: Conformal diagram with the inflationary period

At the top of figure 5.1 the conformal diagram in the Friedmann cosmology is shown. There, the spacetime points p and q were not in causal contact any time before recombination, making the correlations in temperature between the two difficult to explain. Extending the initial singularity to $-\infty$ obviously gets the job done, because now $\eta = 0$ is only the transition time between inflation and the Big Bang. The light cones of p and q can extend infinitely into the conformal past and intersect

somewhere in the diagram. In other words, communication between the two was established during the period of inflation.

Inflation and the flatness problem

Until now a solution was only provided for the horizon problem. In this section we will see that inflation elegantly solves the flatness problem as well. Recall the evolution of the curvature parameter $\Omega_k(t)$ in (5.13), but now normalize it to some initial time t_i (before inflation):

$$\Omega_k(t) = \frac{(a_i H_i)^2}{(aH)^2}\Omega_k(t_i) \tag{5.25}$$

Given the decrease of the Hubble radius during inflation, any curvature will be quickly wiped out, leaving a Universe with zero spatial curvature. To see this more explicitly, consider a non-flat Universe dominated by a fluid with equation of state w. The Friedmann equation reads:

$$H^2 = (a_i H_i)^2 \left[(1 - \Omega_k(t_i)) \left(\frac{a}{a_i} \right)^{-3(1+w)} + \Omega_k(t_i) \right] \tag{5.26}$$

Multiplying by a^2 we get the evolution of the Hubble radius:

$$(aH)^2 = (a_i H_i)^2 \left[(1 - \Omega_k(t_i)) \left(\frac{a}{a_i} \right)^{-(1+3w)} + \Omega_k(t_i) \right] \tag{5.27}$$

Substituting in (5.25) we have:

$$\Omega_k(t) = \frac{\Omega_k(t_i) \left(\frac{a}{a_i} \right)^{(1+3w)}}{(1 - \Omega_k(t_i)) + \Omega_k(t_i) \left(\frac{a}{a_i} \right)^{(1+3w)}} \tag{5.28}$$

The solution has two fixed points, one being $\Omega_k(t_i) = 0$ as we know from before, and the other being $\Omega_k(t_i) = 1$. The qualitative behavior for a fluid with $1 + 3w > 0$, like matter and radiation, is similar: if $\Omega_k(t_i)$ is only slightly different from zero it grows rapidly away with time. For $\Omega_k(t_i) > 0$ the growth slows and converges to the asymptote $\Omega_k = 1$, an empty and Universe with negative curvature. On the other hand, in a positively curved Universe $\Omega_k(t_i) < 0$ the growth accelerates and the curvature diverges at the turnaround point $\dot{a} = 0$, when the Universe starts to fall back in on itself.
The behavior is instead crucially different for an energy density that satisfies $1 + 3w < 0$. In this case, if $\Omega_k(t_i) \neq 0$, the curvature very quickly approaches zero. Therefore the flatness problem is solved if inflation lasts long enough to wipe out any initial curvature. This has to happen to such an extent that the subsequent expansion during the Friedmann cosmology does not increase it above today's observational upper bound $\Omega_k(t_0) < 5 \times 10^{-3}$.

As one final note, notice that inflation does not change the global geometric properties of the Universe. Whether it is open or closed, it will remain so independently

of inflation. What inflation does is it smooths out the local spacetime so that it looks spatially flat with great precision.

The duration of inflation

Before turning to the physical models that realize inflation, let us understand just how much accelerated expansion is needed to solve the problems that arise within the Friedmann cosmology.

Recall that, to solve the horizon problem, the increase of the Hubble radius during matter and radiation domination has to be compensated by a period of inflation where it is decreasing. To provide a lower bound on the duration of inflation, we require that all the fluctuations inside the horizon today were inside the horizon at early times as well. This, we remember, is because the CMB displays a uniform temperature with small correlated perturbations even on the largest scales, i.e. those that have just entered the horizon today $(a_0 H_0)^{-1}$. This condition is expressed as follows:

$$(a_0 H_0)^{-1} < (a_i H_i)^{-1} \tag{5.29}$$

where we denote with i the beginning of inflation. To quantify the amount by which the Hubble sphere decreases during inflation we keep track of the increase of the scale factor in terms of the total number of "e-foldings" (the number of factors of e) between the time when inflation begins (i) and ends (e):

$$N_{\text{tot}} \equiv \log\left(\frac{a_e}{a_i}\right) \tag{5.30}$$

From this perspective, the goal is to find the minimum number of e-foldings that are required to solve the horizon problem. To get a rough estimate, we will assume that the temperature at the end of inflation (the "reheating temperature", we will have more to say on this later) was approximately $T_e \simeq 10^{15}$ GeV, and we will ignore the recent periods of matter and dark energy domination. Since during radiation domination we have $H^2 \propto a^{-2}$ we can write the following:

$$\frac{a_0 H_0}{a_e H_e} = \frac{a_0}{a_e}\left(\frac{a_e}{a_0}\right)^2 = \frac{a_e}{a_0} \sim \frac{T_0}{T_e} \sim 10^{-28}\left(\frac{10^{15}\,\text{GeV}}{T_e}\right) \tag{5.31}$$

where we have used $T_0 \sim 10^{-13}$ GeV and $T(a) \sim 1/a$. Plugging the above relation into the condition (5.29) yields:

$$(a_i H_i)^{-1} > (a_0 H_0)^{-1} \sim 10^{-28}\left(\frac{10^{15}\,\text{GeV}}{T_e}\right)(a_e H_e)^{-1} \tag{5.32}$$

In the next section we will argue that, during inflation, the Hubble parameter is approximately constant. This being the case, we have that $H_e = H_i$ and therefore:

$$N_{\text{tot}} > 64 + \log\left(\frac{T_e}{10^{15}\,\text{GeV}}\right) \tag{5.33}$$

Given the result, we say that the Universe had to expand for roughly more than 60 e-folds for the largest scales to have been in causal contact in the past.

It is also interesting to figure out the number of e-folds necessary to solve the flatness problem. In that case, the condition is that the inflationary period has to reduce the curvature to such an extent as to compensate for its increase during radiation and matter domination. In particular, the curvature today cannot be above the observational bound:

$$|\Omega_k| < 5 \times 10^{-3} \tag{5.34}$$

Recall that this value for the curvature today implies the values (5.14) at the epoch of nucleosynthesis and electroweak phase transition. In the same manner we could extrapolate Ω_k at the start of radiation domination, which corresponds to the end of inflation. Taking the temperature to be $T_e \simeq 10^{15}$ GeV, equation (5.13) gives:

$$\Omega_k(t_e) \simeq 10^{-56} \tag{5.35}$$

Finally, recalling (5.25) we have that:

$$\frac{\Omega_k(t_e)}{\Omega_k(t_i)} = \frac{(a_i H_i)^2}{(a_e H_e)^2} = \left(\frac{a_i}{a_e}\right)^2 = e^{-2N_{\text{tot}}} \tag{5.36}$$

where we have used $H_e = H_i$. Again, to get a rough estimate of the number of e-folds, taking $\Omega_k(t_i)$ of order unity we get:

$$N_{\text{tot}} \simeq 64 \tag{5.37}$$

In other words, it is enough to require 64 e-folds to solve the flatness problem. Obviously, if inflation lasted for even more e-folds the curvature today would be very close to zero with great precision. One can then say that a generic prediction of inflation is that, at present time:

$$\Omega_k = 1 \iff \text{INFLATION} \tag{5.38}$$

5.4 The physics of inflation

The advantages of a period of accelerated expansion, or equivalently of a shrinking Hubble sphere, should now be clear enough. However, we have yet to propose a physical candidate that realizes inflation. As far as we understand it, this candidate has to obey the following conditions:

(i) Its equation of state w has to satisfy $w < -1/3$ as required by the second Friedmann equation;

(ii) The period of inflation it gives rise to has to last for a sufficiently long time in order to solve both the horizon and flatness problems. As we laid out, roughly 60 e-folds are sufficient;

(iii) It should be dynamic enough such that inflation ends successfully and continuously connects to the Big Bang (the beginning of radiation domination).

A real scalar field

Let us focus for now on condition (i). This condition can be attained by means of a simple real scalar field, which has been dubbed the "inflaton" in modern literature. The action of a scalar field in curved spacetime reads:

$$S = \int d^4x \sqrt{-g}\mathcal{L} = \int d^4x \sqrt{-g}\left[-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\right] \tag{5.39}$$

where $\sqrt{-g} = a^3$ for the FRW metric and $V(\phi)$ is the inflaton's potential which we leave unspecified for the time being. Solving the Euler-Lagrange equations

$$\partial^\mu \frac{\delta(\sqrt{-g}\mathcal{L})}{\delta(\partial^\mu\phi)} - \frac{\delta(\sqrt{-g}\mathcal{L})}{\delta\phi} = 0 \tag{5.40}$$

we obtain:

$$\ddot{\phi} + 3H\dot{\phi} - \frac{\nabla^2\phi}{a^2} + \frac{dV}{d\phi} = 0 \tag{5.41}$$

Interesting is the appearance of the friction term $3H\dot{\phi}$ due to the expansion of the Universe. The field, as it moves in its potential, suffers a friction that is not present in a static spacetime. The energy-momentum tensor of a scalar field reads:

$$T^\mu_\nu = \partial^\mu\phi\partial_\nu\phi - \delta^\mu_\nu\left[\frac{1}{2}\partial^\alpha\phi\partial_\alpha\phi + V(\phi)\right] \tag{5.42}$$

The time component T^0_0 is equal to $-\rho_\phi$ and the pressure is $P_\phi = 1/3T^i_i$:

$$\begin{aligned}\rho_\phi &= \frac{\dot{\phi}^2}{2} + V(\phi) + \frac{(\nabla\phi)^2}{2a^2}\\ P_\phi &= \frac{\dot{\phi}^2}{2} - V(\phi) - \frac{(\nabla\phi)^2}{6a^2}\end{aligned} \tag{5.43}$$

If the gradient term $\nabla\phi$ were dominant we would obtain $P_\phi = -\rho_\phi/3$, which is not enough to power inflation. We will assume that the field is homogeneous to zeroth order, consisting of a zeroth order infinite wavelength part and a perturbation which depends on the position:

$$\phi(x^\mu) = \phi_0(t) + \delta\phi(t, \vec{x}) \tag{5.44}$$

In this section we will be concerned with the homogeneous part of the field, justified by the fact that the fluctuations $\delta\phi$ are much smaller. This way we can neglect the gradient term and write the equation of state of the field as (we set $c = 1$ from here on):

$$w_\phi = \frac{P_\phi}{\rho_\phi} = \frac{\frac{\dot{\phi}_0^2}{2} - V(\phi_0)}{\frac{\dot{\phi}_0^2}{2} + V(\phi_0)} \tag{5.45}$$

Crucially, if we impose the condition:

$$V(\phi_0) \gg \dot{\phi}_0^2 \tag{5.46}$$

we obtain:

$$P_\phi = -\rho_\phi \iff w_\phi = -1 \tag{5.47}$$

From this straightforward calculation we realize that a configuration of a scalar field whose energy dominates the Universe and whose potential dominates the kinetic term can drive inflation.

A slowly rolling field

Let us now dwell on condition (ii), which requires inflation to last long enough. To quantify this idea, we will introduce the so-called "slow-roll parameters" and impose that condition on them. A key characteristic of inflation is that physical quantities are slowly varying, despite the very rapid expansion of space. We remember the inflationary condition to be:

$$\frac{d}{dt}(aH)^{-1} = -\frac{\dot{a}H + a\dot{H}}{(aH)^2} \equiv -\frac{1}{a}(1+\epsilon) \tag{5.48}$$

where we have introduced the slow-roll parameter ϵ:

$$\epsilon = -\frac{\dot{H}}{H^2} = -\frac{d\log H}{dN} \tag{5.49}$$

where $dN = d\log a = Hdt$. Since the Hubble radius shrinks, this parameter has to satisfy $\epsilon < 1$, meaning that the fractional change in Hubble parameter per e-fold of expansion is small. It can be showed that the scale invariance of primordial fluctuations generated by inflation actually requires $\epsilon \ll 1$, implying that $H = \text{const.}$ during inflation, a fact we used in the previous section.
Since it is also necessary for inflation to last sufficiently long, ϵ has to be small for a sufficiently large number of e-foldings. This condition is measured by a second slow-roll parameter κ:

$$\kappa \equiv \frac{d\log\epsilon}{dN} = \frac{\dot{\epsilon}}{H\epsilon} \tag{5.50}$$

Since we need $\epsilon \ll 1$ to last, we require also that $|\kappa| \ll 1$. These conditions can be imposed directly on the field in the following way. The inflaton being the dominating energy, the complete action during inflation is:

$$S = \int d^4x\sqrt{-g}\left(R - \frac{1}{2}\partial^\mu\phi_0\partial_\mu\phi_0 - V(\phi_0)\right) \tag{5.51}$$

Varying with respect to $g_{\mu\nu}$ and ϕ_0 yields the equations that govern the dynamics of the Universe:

$$\begin{aligned} H^2 &= \frac{1}{3M_{\rm Pl}^2}\left(\frac{1}{2}\dot{\phi}_0^2 + V(\phi)\right) \\ \ddot{\phi}_0 + 3H\dot{\phi}_0 &= -\frac{dV(\phi_0)}{dt} \end{aligned} \tag{5.52}$$

where we have introduced the reduced Planck mass $M_{\rm Pl}^2 = (8\pi G)^{-1}$. We immediately note that the two equations can be combined to find the evolution of the Hubble parameter:

$$\dot{H} = -\frac{1}{2}\frac{\dot{\phi}_0^2}{M_{\rm Pl}^2} \tag{5.53}$$

Taking the ratio of the Friedmann equation (5.52) and (5.53) yields an expression for the slow roll parameter ϵ:

$$\epsilon = -\frac{\dot{H}}{H^2} = \frac{1/2\dot{\phi}_0^2}{M_{\rm Pl}H^2} = \frac{3/2\dot{\phi}_0^2}{1/2\dot{\phi}_0^2 + V(\phi_0)} \tag{5.54}$$

Inflation therefore occurs if $\dot{\phi}_0^2 \ll V(\phi_0)$, a condition we already found before. Due to the kinetic energy of the field being small, this situation is called "slow-roll inflation", and sees the inflaton slowly-rolling on the potential. As explained, in order for this slow-roll to persist, we require that $|\kappa| \ll 1$. This is actually equivalent to imposing the inflaton's acceleration to be small too, as we can see by introducing a new slow-roll parameter δ:

$$\delta \equiv -\frac{\ddot{\phi}_0}{H\dot{\phi}_0} \tag{5.55}$$

Taking the time derivative of (5.54) we get:

$$\dot{\epsilon} = -\frac{\dot{\phi}_0\ddot{\phi}_0}{M_{\rm Pl}^2H^2} - \frac{\dot{\phi}_0^2\dot{H}}{M_{\rm Pl}^2H^3} \tag{5.56}$$

Plugging (5.56) into the definition of κ (5.50):

$$\kappa = \frac{\dot{\epsilon}}{H\epsilon} = 2\frac{\ddot{\phi}_0}{H\dot{\phi}_0} - 2\frac{\dot{H}}{H^2} = 2(\epsilon - \delta) \tag{5.57}$$

Therefore we see that $\{\epsilon, |\kappa|\} \ll 1$ implies $\{\epsilon, |\delta|\} \ll 1$ so that the two sets of parameters are interchangeable. This line of reasoning shows that if both the speed and the acceleration of the inflaton field are small, then the accelerated expansion will continue for a long time. This is slow-roll inflation.
The conditions for prolonged inflation can equivalently be cast on the form of the potential in the following way. First, note that the condition $\epsilon \ll 1$, or $\dot{\phi}_0^2 \ll V(\phi_0)$, implies the simplification of the Friedmann equation in (5.52):

$$H^2 \simeq \frac{V(\phi_0)}{3M_{\rm Pl}^2} \tag{5.58}$$

The Hubble expansion rate is therefore strictly determined by the inflaton's potential V. Then, the condition $|\delta| \ll 1$ simplifies the inflaton's dynamics:

$$3H\dot{\phi}_0 \simeq -\frac{dV(\phi_0)}{d\phi_0} \tag{5.59}$$

This equation provides a simple relation between the slope of the potential and the field's kinetic energy, such that requiring a small kinetic energy is equivalent to requiring some degree of flatness in the potential. Plugging both simplifications (5.58) and (5.59) into (5.54) yields:

$$\epsilon = \frac{1/2\dot{\phi}_0^2}{M_{\rm Pl}^2H^2} \simeq \frac{M_{\rm Pl}^2}{2}\left(\frac{V_{,\phi_0}}{V}\right)^2 \equiv \epsilon_V \tag{5.60}$$

where we use the subscript V when a certain parameter is specified in terms of the potential. The second condition on the potential is conveniently found by combining the parameters ϵ and δ. First take the derivative of the second equation in (5.52):

$$3\dot{H}\dot{\phi}_0 + 3H\ddot{\phi}_0 = -V_{,\phi_0\phi_0}\dot{\phi}_0 \tag{5.61}$$

Now define the new parameter η_V as:

$$\delta + \epsilon = -\frac{\ddot{\phi}_0}{H\dot{\phi}} - \frac{\dot{H}}{H^2} \simeq M_{\rm Pl}^2 \frac{V_{,\phi_0\phi_0}}{V} \equiv \eta_V \tag{5.62}$$

Hence, a convenient and straightforward way to judge whether a given potential $V(\phi_0)$ leads to slow-roll inflation is to check that $\{\epsilon_V, \eta_V\} \ll 1$.
Summarizing this discussion, a period of inflation driven by a real scalar field is obtained if the following field-slow-roll parameters are small:

$$\begin{aligned} \epsilon &= \frac{3/2\dot{\phi}_0^2}{1/2\dot{\phi}_0^2 + V(\phi_0)} \\ \delta &= -\frac{\ddot{\phi}_0}{H\dot{\phi}_0} \end{aligned} \tag{5.63}$$

Equivalently, the potential-slow-roll parameters have to be small:

$$\begin{aligned} \epsilon_V &= \frac{M_{\rm Pl}^2}{2}\left(\frac{V_{,\phi_0}}{V}\right)^2 \\ \eta_V &= M_{\rm Pl}^2 \frac{V_{,\phi_0\phi_0}}{V} \end{aligned} \tag{5.64}$$

The prototype for the potential $V(\phi_0)$ is plotted in the following figure:

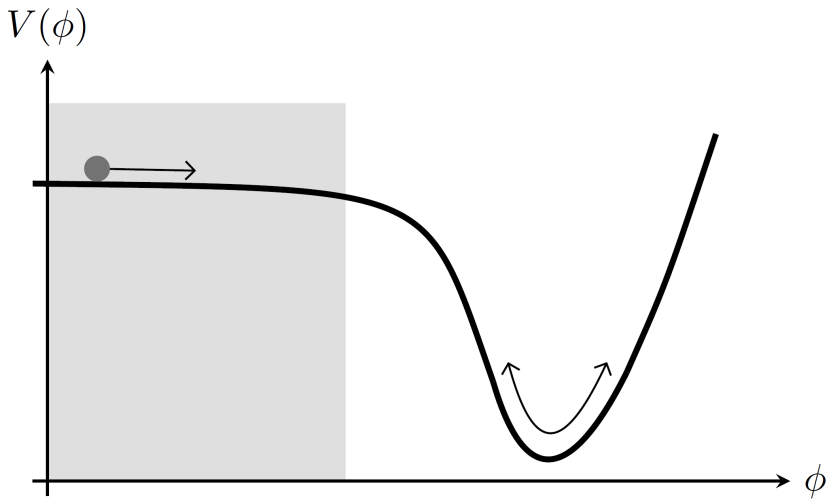

FIGURE 5.5: Form of the potential $V(\phi_0)$: slow-roll is obtained on top of the hill

As long as the field is slowly rolling on top of the hill (the shaded region), the physical quantities are slowly varying and accelerated expansion is sustained.

As a final application of inflation, we note that within the slow-roll approximations it is easy to compute the the number of e-foldings between the start and end of inflation. Denoting by $\phi_{0,i}$ and $\phi_{0,e}$ the values of the inflaton at the beginning and end of inflation, N_{tot} is calculated as:

$$\begin{aligned} N &= \int_{t_i}^{t_e} H dt \simeq \int_{\phi_{0,i}}^{\phi_{0,e}} d\phi_0 \, \frac{H}{\dot{\phi}_0} \simeq -3 \int_{\phi_{0,i}}^{\phi_{0,e}} d\phi_0 \, \frac{H^2}{V'} \\ &= -\frac{1}{M_{\text{Pl}}^2} \int_{\phi_{0,i}}^{\phi_{0,e}} d\phi_0 \, \frac{V}{V'} \end{aligned} \tag{5.65}$$

The field values $\phi_{0,i}$ and $\phi_{0,e}$ are calculated at the boundary of the interval where $\epsilon_V < 1$, in the sense that the beginning and end of inflation are defined by $\epsilon(t_i) = \epsilon(t_e) = 1$, which implies that the accelerated expansion is over:

$$\epsilon = -\frac{\dot{H}}{H^2} = 1 - \frac{\ddot{a}}{H^2 a^2} = 1 \iff \ddot{a} = 0 \tag{5.66}$$

Since the first in (5.63) relates ϵ to the inflaton, we see that $\epsilon = 1$ is satisfied if $\dot{\phi}_0^2 = V(\phi_0)$, that is the field starts to acquire enough kinetic energy such that it becomes of the same order of the potential energy. This happens, for instance, when the field starts rolling down the hill in figure (5.5). This discussion segways nicely into the final section about the complex topic of reheating.

The end of inflation and reheating

We now finally turn to the condition (iii) which regards the end of the inflationary era.

The inflaton, being a scalar field, is a dynamic enough object that it accommodates neatly for the end of accelerated expansion. In fact, as was explained just above, inflation ends when the energy associated with the inflaton becomes smaller than the kinetic energy of the field. Usually this happens near the minimum of the potential, as in figure (5.5).

However, at the end of inflation the Universe is typically in a highly non-thermal state, due to inflation's ability to homogeneise the Universe and thus to leave it at effectively zero temperature. For this reason, any successful theory of inflation must explain not only how inflation ends, but how the cosmos was reheated to the high temperatures required for the radiation dominated epoch to begin. At the very least primordial nucleosynthesis requires the Universe to be close to thermal equilibrium at a temperature around 1 MeV. Remarkably, the scalar field framework can also account for this phenomenon. Since the theories complicate substantially, in this section we will only illustrate the main ideas.

The "old" theory of reheating, developed shortly after the first inflationary theories, was based on single particle decays. The energy density of the inflaton is converted to ordinary particles which quickly thermalize. Since the said energy density was the same everywhere (except for small perturbations) in the Universe, the temperature was likewise uniform. This reheating temperature T_{RH} can be calculated by assuming an instantaneous conversion of the inflaton's energy density into radiation, when the decay width of the inflaton Γ_ϕ is equal to H (otherwise the expansion prevents the decay). As we said, the inflaton field, when inflation ends, is near the minimum of the potential at some $\phi_0 \simeq \phi_m$. Here, it executes oscillations and the

potential can be approximated by:

$$V(\phi_0) \simeq \frac{1}{2}V''(\phi_m)(\phi_0 - \phi_m)^2 \equiv \frac{1}{2}m^2(\phi_0 - \phi_m)^2 \tag{5.67}$$

The equation of motion for ϕ_0 becomes:

$$\ddot{\phi}_0 + 3H\dot{\phi}_0 + m^2(\phi_0 - \phi_m) = 0 \tag{5.68}$$

The solution is:

$$\phi_0(t) = \phi_i \left(\frac{a_i}{a}\right)^3 \cos(m(t - t_i)) \tag{5.69}$$

where now the subscript i denotes the beginning of oscillations. The equation satisfied by the energy density in the oscillating field over many oscillations is:

$$\begin{aligned} \langle \dot{\rho}_\phi \rangle &= \left\langle \frac{d}{dt}\left(\frac{1}{2}\dot{\phi}_0 + V(\phi_0)\right)\right\rangle = \langle \dot{\phi}_0 \left(\ddot{\phi}_0 + V'(\phi_0)\right)\rangle \\ &= \langle \dot{\phi}_0 \left(-3H\dot{\phi}_0\right)\rangle = -3H\left\langle \dot{\phi}_0^2 \right\rangle = -3H\left\langle \rho_\phi \right\rangle \end{aligned} \tag{5.70}$$

where we used the equipartition of the energy density during oscillations $\langle \dot{\phi}_0^2/2\rangle = \langle V(\phi_0)\rangle = \langle \rho_\phi/2\rangle$. The solution is (removing averaging symbols):

$$\rho_\phi = (\rho_\phi)_i \left(\frac{a_i}{a}\right)^3 \tag{5.71}$$

The Hubble parameter is then:

$$H^2 = \frac{1}{M_{\rm Pl}^2}\rho_\phi = \frac{(\rho_\phi)_i}{M_{\rm Pl}^2}\left(\frac{a_i}{a}\right)^3 \tag{5.72}$$

Equating H to Γ_ϕ allows to find an expression for $(a_{\rm RH}/a_i)$, and therefore for the reheating temperature. As we argued, we could simplify matters by assuming that all the available energy density is instantaneously converted into radiation at the time $a_{\rm RH}$. Follwing this logic, we set ρ_ϕ equal to the radiation energy density:

$$\rho_r = \frac{\pi^2}{30} g_\star T_{\rm RH}^4 \tag{5.73}$$

where $g_\star$ is the effective number of relativistic degrees of freedom at the temperature $T_{\rm RH}$ (see the theory on Big Bang nucleosynthesis). Equating H to Γ_ϕ results in:

$$T_{\rm RH} = \left(\frac{90}{g_\star \pi^2}\right)^{1/4} \sqrt{\Gamma_\phi M_{\rm Pl}} \tag{5.74}$$

The inflaton and large-scale structure

The realization of inflation via a real scalar field is a remarkable framework. Not only does it solve elegantly the horizon and flatness problems, but it is also dynamical enough to last a long time and end by reheating the Universe, transitioning continuously into the radiation dominated epoch.
However, the inflaton also realizes a natural mechanism that gives rise to the deviations from inhomogeneity that are observed in the Universe on scales $\lesssim$ 100 Mpc

(even intuitively, at a local level galaxies are distributed far from homogeneously). Remember the inflaton's decomposition:

$$\phi(x^\mu) = \phi_0 + \delta\phi(t, \vec{x}) \tag{5.75}$$

It turns out that treating the perturbations $\delta\phi$ quantum mechanically sets the initial conditions for the Universe: before inflation it is riddled with inhomogeneities due to the uncertainty principle. These are pushed outside the horizon during inflation and start to grow only when they re-enter during radiation and matter domination, when finally they become the large scale structures that are observed today in the Universe. Inflation therefore provides a beautiful and natural link between the quantum mechanical microphysics and the macrophysics of the Universe. This is one of the few synergies that have been found between high energy physics and gravity.

Part 2: The Inhomogeneous Universe

Chapter 6

The Evolution of Matter and Radiation in a Perturbed Universe

In the previous section, making use of large scales observations, we have treated the Universe as perfectly homogeneous and isotropic. However, as was pointed out above, the presence of structures in the Universe warrants the introduction of inhomogeneities in the metric and in the matter-energy content, whose evolution, as long as their values are small, can be followed with linear perturbation theory.
After neutrino decoupling and BBN, the only relevant forces that play a critical role in cosmological evolution are gravity, which is the same for all particles due to the equivalence principle, and electromagnetism. Before recombination, photons, electrons, protons and neutrons are tightly coupled by their electromagnetic interactions and are thus treated usually as a single fluid. The two main components that drive the evolution of the Universe are the homogeneous densities of the species, which govern the cosmological expansion rate, and their perturbations, which source the local gravitational potential. This physical process is, in general, highly non-linear, and it can lead to very non-trivial dynamics. In practice, one has to solve coupled differential equations for the metric and the matter-energy perturbations. These equations are given by the Einstein equations and the Boltzmann equations.
The two important times in the evolution of the Universe are matter-radiation equality, at around $z_{\rm eq} \sim 3400$, and photon decoupling at $z_{\rm dec} \sim 1100$ (which basically coincides with recombination). They are reference points because, before equality, the growth of density perturbations is damped by the high pressure exerted by photons, therefore the growth only starts to become significant when the Universe becomes matter dominated. Moreover, before decoupling, the baryons are coupled to photons and their density perturbations oscillate, as opposed to the dark matter ones, which decouples very early on. These oscillations, sound waves in the primordial plasma, arise as a result of the dance between two opposing forces: the clustering effect of gravity battles with the radiation pressure which causes the perturbations to re-expand; the frequency of these oscillations depends on the wavelength of the fluctuations, so that at decoupling modes with different wavelengths are captured at different phases in their evolutions, giving rise to the characteristic CMB patterns. After decoupling, when the electrons get captured by the protons to make the first atoms, the baryons lose the pressure support of the photons and their perturbations begin to grow, and eventually they fall into the potential wells formed by the already clumped dark matter. This is the start of structure formation.

6.1 The Boltzmann Equation Formalism

The aim of this chapter is to understand how matter, photons and neutrinos behave in a given expanding and perturbed spacetime. In principle, the evolution of the dark energy component, and its perturbations, should be followed as well. The perturbations however, exist only if dark energy is not a cosmological constant, so for the purposes of this chapter it will be reasonable to exclude them.
While in the homogeneous and isotropic Universe, the matter-energy content was treated as a perfect fluid, as justified by the symmetries, here the correct course of action is to make use of the Boltzmann equation formalism, which is concerned with the evolution of the distribution function $f(\mathbf{x}, \mathbf{p}, t)$, for which the master equation is:

$$\frac{df_s}{dt} = C[f_s] \tag{6.1}$$

where the term on the right is related to the interactions regarding the species of particle considered, and is called the "collision term". The subscript s stands to label the spin of the particles. This formalism includes the case in which the particles behave, on large scales, as a perfect fluid, but it also extends it to the description of cases where the behavior deviates radically from that of a fluid. For example, before recombination, baryons and photons can be treated as fluids, as said above, because of the high interaction rate. Indeed, the Boltzmann equations governing these components during this era can also be derived using the perfect fluid formalism (with a suitable choice of perturbations). However, after recombination the photons free stream and therefore don't behave as a fluid. Then, if one is to calculate the CMB anisotropies, the only possible method is to follow the distribution function itself, that is to use the Boltzmann equation.
The distribution function allows us to derive all macroscopic properties of the collection of particles that is being considered, such as the density, pressure and velocities. Particular importance is given to the energy-momentum tensor, which can be found by:

$$T^{\mu}_{\nu}(\mathbf{x}, \mathbf{p}) = \frac{g_s}{\sqrt{-\det[g_{\mu\nu}]}} \int \frac{d^3P}{(2\pi)^3} \frac{P^{\mu}P_{\nu}}{P^0} f_s(\mathbf{x}, \mathbf{p}, t) \tag{6.2}$$

where g_s is the degeneracy factor for spins, and $P^{\mu} = dx^{\mu}/d\lambda$ is the comoving momentum. The prefactor involving the determinant of the metric is necessary in order to ensure that T^{μ}_{ν} obeys the correct conservation law $\nabla_{\mu} T^{\mu}_{\nu} = 0$. As anticipated, it is easy to check that this form reduces to the form of the energy-momentum tensor introduced in (3.8) for the FRW metric. Equation (6.2) is also a neat way of organizing the expressions in (3.20).
A fundamental concept to keep in mind when treating perturbations, both in the metric and in the energy-momentum, is that they will always be approximated to linear order, that is all terms quadratic or higher orders in them will be neglected. This approximation holds true for perturbations in the radiation throughout the whole history of the Universe, since they were frozen at recombination, when they were indeed small in value. On the other hand, the dark matter perturbations, not being affected by interactions, start to grow early on and reach values that linear perturbation theory is not able to describe. Following the matter perturbations in this stage requires a more complex formalism of structure formation.

The Boltzmann equation in an expanding spacetime

To begin, one must specify the form of the metric that accounts for small perturbations around the FRW spacetime. For the moment, we introduce the following form:

$$ds^2 = -(1+\Psi(\mathbf{x},t))dt^2 + a^2(t)(1+\Phi(\mathbf{x},t))d\mathbf{x}^2 \tag{6.3}$$

Here, for simplification purposes, the flat form of the FRW metric was used, which is recovered as a zero-th order expansion in the perturbation variables Ψ and Φ. Additionally, a specific gauge was chosen, called the "Newtonian gauge", which will be explained in more detail below. The perturbation Ψ corresponds to the Newtonian potential and governs the motion of non-relativistic bodies, whereas the variable Φ can be thought of as a local perturbation in the scale factor, meaning that $a(t) \to a(\mathbf{x},t) = a(t)\sqrt{1+\Phi(\mathbf{x},t)}$.
The calculation of the right hand side of the Boltzmann equation (6.1) requires the knowledge of how particles move in an expanding spacetime, since the distribution function f_s depends on the position $\mathbf{x}$ and the momentum $\mathbf{p}$, such that the total derivative with respect to time will include terms like $d\mathbf{x}/dt$ and $d\mathbf{p}/dt$. These are found via the geodesic equation associated to the metric (6.3), whose Chrsitoffel's symbols read, to first order:

$$\begin{aligned}
\Gamma^0_{00} &= \dot{\Psi} \\
\Gamma^0_{0i} &= \partial_i \Psi \\
\Gamma^0_{ij} &= \delta_{ij} a^2 [H + 2H(\Phi - \Psi) + \dot{\Phi}]
\end{aligned} \tag{6.4}$$

and

$$\begin{aligned}
\Gamma^i_{00} &= \frac{1}{a^2}\partial_i \Psi \\
\Gamma^i_{0j} &= \delta_{ij}(H + \dot{\Phi}) \\
\Gamma^i_{jk} &= [\delta_{ij}\partial_k + \delta_{jk}\partial_i]\Phi
\end{aligned} \tag{6.5}$$

Note that by convention, both δ_{ij} and spatial derivatives δ_i live in Euclidian space, so that their upper and lower indices can be interchanged freely. Next, we note that the mass-shell constraint for a particle with mass m is given by:

$$g_{\mu\nu}P^\mu P^\nu = -(1+2\Psi)(P^0)^2 + p^2 = -m^2 \tag{6.6}$$

where the physical momentum is defined as above:

$$p^2 = g_{ij}P^i P^j \tag{6.7}$$

and the energy is defined by $E^2 = p^2 + m^2$. The mass-shell constraint allows us to rewrite P^0 and P^i in favor of E and p, such that:

$$P^\mu = \left(E(1-\Psi), p^i \frac{1-\Phi}{a}\right) \tag{6.8}$$

where it was convenient to split the physical momentum into p and $\hat{p}^i$:

$$p^i = p\hat{p}^i \tag{6.9}$$

and $\hat{p}^i$ is a unit vector, that is $\delta_{ij}\hat{p}^i\hat{p}^j = 1$. Plugging (6.8) into the expression for the energy-momentum tensor (6.3) will be the method used to compute the right hand side of the Einstein equations.
At linear order one finds:

$$\frac{dx^i}{dt} = \frac{dx^i}{d\lambda}\frac{d\lambda}{dt} = \frac{P^i}{P^0} = \frac{\hat{p}^i}{a}\frac{p}{E}(1 - \Phi + \Psi) \tag{6.10}$$

The remaining term to be calculated is dp^i/dt, from which dp/dt and $d\hat{p}^i/dt$ can be later deduced. First, the derivative of p^i along a geodesic reads:

$$\frac{dp^i}{dt} = P^i\frac{d}{d\lambda}[(1+\Phi)a] + (1+\Phi)a\frac{dP^i}{d\lambda} \tag{6.11}$$

The first term can be computed using $d/d\lambda = P^\mu\partial/\partial^\mu$:

$$\frac{d}{d\lambda}[(1+\Phi)a] = P^0 a[H + \dot{\Phi}] + aP^k\partial_k\Phi \tag{6.12}$$

whereas the second term is recovered through the geodesic equation:

$$\begin{aligned}\frac{dP^i}{d\lambda} &= -\Gamma^i_{\alpha\beta}P^\alpha P^\beta \\ &= -E\left(\frac{E}{a^2}\partial_i\Psi + 2(H+\dot{\Phi})\frac{p^i}{a}(1-\Psi-\Phi) + \frac{2}{a^2}\frac{p^i}{E}p^k\partial_k\Phi - \frac{p^2}{a^2E}\partial_i\Phi\right)\end{aligned} \tag{6.13}$$

With these two expressions, equation (6.11) becomes:

$$\begin{aligned}\frac{dp^i}{d\lambda} &= E(1-\Psi)\left((H+\dot{\Phi})p^i + p^k\partial_k\Phi\frac{p^i}{aE}\right) \\ &\quad - E\left(\frac{E}{a}\partial_i\Psi + 2(H+\dot{\Phi})p^i(1-\Psi) + \frac{2}{a}\frac{p^i}{E}p^k\partial_k\Phi - \frac{p^2}{aE}\partial_i\Phi\right)\end{aligned} \tag{6.14}$$

Finally, the derivative with respect to λ can be converted into one with respect to time using the definition of P^0:

$$\begin{aligned}\frac{dp^i}{dt} &= [H+\dot{\Phi}]p^i + p^k\partial_k\frac{p^i}{aE} \\ &\quad - \left(\frac{E}{a}\partial_i\Psi + 2(H+\dot{\Phi})p^i + \frac{2}{a}\frac{p^i}{E}p^k\partial_k\Phi - \frac{p^2}{aE}\partial_i\Phi\right)\end{aligned} \tag{6.15}$$

This expression simplifies to:

$$\frac{dp^i}{dt} = -(H+\dot{\Phi})p^i - \frac{E}{a}\partial_i\Psi - \frac{1}{a}\frac{p^i}{E}p^k\partial_k\Phi + \frac{p^2}{aE}\partial_i\Phi \tag{6.16}$$

Using the definition of the physical momentum, one has:

$$\frac{dp}{dt} = \frac{d}{dt}\sqrt{\delta_{ij}P^iP^j} = \delta_{ij}\frac{p^i}{p}\frac{dp^j}{dt} \tag{6.17}$$

for the absolute value of p^i, so that finally:

$$\frac{dp}{dt} = -[H + \dot{\Phi}]p - \frac{E}{a}\hat{p}^i\partial_i\Psi \tag{6.18}$$

The form of (6.18) can be understood as follows. The first term corresponds to the loss of momentum due to the usual friction caused by the expansion of the Universe, where the term $\dot{\Phi}$ is expected from the interpretation of Φ as a perturbation in $a(t)$. The last term, on the other hand, describes the effect of a particle that travels in ($\hat{p}^i\partial_i\Psi < 0$) or out ($\hat{p}^i\partial_i\Psi > 0$) of wells gains or loses energy respectively. Note that while the terms $\partial_i\Phi$ and $p^k\partial_k\Phi$ disappear when going from dp^i/dt to dp/dt, they do affect the direction of momentum at linear order:

$$\frac{d\hat{p}^i}{dt} = \frac{1}{p}\frac{dp^i}{dt} - \frac{p^i}{p^2}\frac{dp}{dt} = \frac{E}{ap}(\delta^{ik} - \hat{p}^i\hat{p}^k)\partial_k\left(\frac{p^2}{E^2}\Phi - \Psi\right) \tag{6.19}$$

Therefore, spatial gradients of the potentials change the direction of both massless and massive particles. The result (6.16) correctly reproduces the Newtonian result $d\mathbf{p}/dt = -m\nabla\Psi$, since $p \ll E \sim m$, except for a factor of a^{-1}, which is however straightforward when considering the relation between comoving and physical coordinates.

The collisionless Boltzmann equation for massless particles

The collisionless Boltzmann equation for ultra relativistic particles is, setting $C[f] = 0$ in (6.1) and expanding the derivative:

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i}\frac{\partial x^i}{\partial t} + \frac{\partial f}{\partial p}\frac{\partial p}{\partial t} + \frac{\partial f}{\partial \hat{p}^i}\frac{\partial \hat{p}^i}{\partial t} = 0 \tag{6.20}$$

We then specialize to the case $m = 0$, i.e. $E = p$ and use the derivatives with respect to time calculated in (6.10), (6.18) and (6.19):

$$\begin{aligned} 0 = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i}\frac{\hat{p}^i}{a}(1 - \Phi + \Psi) - \frac{\partial f}{\partial p}\left([H + \dot{\Phi}]p + \frac{1}{a}p^i\partial_i\Psi\right) \\ + \frac{\partial f}{\partial \hat{p}^i}\frac{1}{a}\left(\partial_i(\Phi - \Psi) - \hat{p}^i\hat{p}^k\partial_k(\Phi - \Psi)\right) \end{aligned} \tag{6.21}$$

In practice, this equation can be simplified by making use of the fact that the zero-th order distribution function, which is in this case of the Bose-Einstein form, does not depend either on $\mathbf{x}$ or the direction $\hat{\mathbf{p}}$ because of homogeneity and isotropy. In addition, working at first order in the perturbations, one now assumes that the order of the perturbations in the metric are of the same order of the deviations from the equilibrium distribution. While, for now, this assumption only simplifies the calculations, it will be verified only later. Working in this framework, then, the last term can be immediately dropped, given that $\frac{\partial f}{\partial \hat{p}^i}$ is already a first order term, which multiplies the potentials in the bracket. Furthermore, the potentials in the second term can be neglected for the same reason, yielding the simplified form:

$$\frac{\partial f}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} - \left(H + \dot{\Phi} + \frac{1}{a}\hat{p}^i\partial_i\Psi\right)p\frac{\partial f}{\partial p} = 0 \tag{6.22}$$

This is the master equation that describes the propagation of photons, and therefore the CMB anisotropies.

The collisionless Boltzmann equation for massive particles

The analogous derivation can be carried out for massive particles, which brings (6.20) to:

$$\frac{\partial f}{\partial t} + \frac{p}{E}\frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} - \left(H + \dot{\Phi} + \frac{E}{ap}\hat{p}^i\partial_i\Psi\right) p\frac{\partial f}{\partial p} = 0 \tag{6.23}$$

This is obviously the same as (6.22), except for the factors of p/E, that in the massless case vanish.

Collision terms

The term on the right hand side of the Boltzmann equation, $C[f]$, is concerned with the interactions between particles, as stated above. By interactions, one usually refers to a wide variety of processes, which include scattering as well as pair creation, annihilation, and particle decay. A general process where particles of type 1 and 2 interact to form particles 3 and 4 is schematically written as:

$$(1)_{\mathbf{p}} + (2)_{\mathbf{q}} \longleftrightarrow (3)_{\mathbf{p}'} + (4)_{\mathbf{q}'} \tag{6.24}$$

where the subscripts indicate the momenta. By conservation of four-momentum, physical momenta and energy are separately conserved:

$$\mathbf{p} + \mathbf{q} = \mathbf{p}' + \mathbf{q}' \quad E_1(\mathbf{p}) + E_2(\mathbf{q}) = E_3(\mathbf{p}') + E_4(\mathbf{q}') \tag{6.25}$$

where, again, $E(p) = \sqrt{p^2 + m^2}$ and p is the physical momentum as defined by (6.7), which also appears as an variable in the distribution function $f_s = f_s(\mathbf{x}, \mathbf{p}, t)$. Here, the subscript s indicates the spin, i.e. the spin degeneracy g_s, and $s = 1, 2, 3, 4$. The reaction (6.24) affects the distribution function f_s of, say, particle of type 1, by taking away those type 1 particles that get scattered away from momentum $\mathbf{p}$ by the forward reaction, and by adding those that get scattered into momentum $\mathbf{p}$ by the reverse reaction. Therefore, we must sum over all other momenta $(\mathbf{q}, \mathbf{p}', \mathbf{q}')$ (that satisfy momentum conservation):

$$\begin{aligned} C[f_1(\mathbf{p})] = &\sum_{\mathbf{q},\mathbf{p}',\mathbf{q}'}^{\mathbf{p}+\mathbf{q}=\mathbf{p}'+\mathbf{q}'} \delta_D^{(1)}(E_1(p) + E_2(q) - E_3(p') - E_4(q'))\, |\mathcal{M}|^2 \\ &\times [f_3(\mathbf{p}')f_4(\mathbf{q}') - f_1(\mathbf{p})f_2(\mathbf{q})] \end{aligned} \tag{6.26}$$

where the delta function is there to enforce energy conservation. Note that the arguments $(\mathbf{x}, t)$ do not appear, since all the distribution functions that appear in (6.26) are evaluated at the same point in spacetime. The scattering amplitude $\mathcal{M}$ can be calculated with Quantum Field Theory and depends on the specific process at hand. Technically, one also needs to introduce in the collision term above also quantum effects such as stimulated emission and the Pauli exclusion principle, which amounts to adding factors of $(1 \pm f_3)(1 \pm f_4)$ to the forward reaction, and $(1 \pm f_1)(1 \pm f_2)$ to the reverse reaction. In each case, a plus sign is assigned when the corresponding particle is a boson, and viceversa if it is a fermion. Finally, to perform the sums over

phase space, one needs to take account of the fact that 4-dimensional phase space integrals become 3-dimensional, since the energy is constrained by the on-shell conditions. Therefore, for the collision term affecting particle of type 1, one should write:

$$\begin{aligned} C[f_1(\mathbf{p})] = &\frac{1}{2E_1(p)} \int \frac{d^3q}{(2\pi)^3 2E_2(q)} \int \frac{d^3p'}{(2\pi)^3 2E_3(p')} \int \frac{d^3q'}{(2\pi)^3 2E_4(q')} |\mathcal{M}|^2 \\ &\times (2\pi)^4 \, \delta_D^{(3)}[\mathbf{p}+\mathbf{q}-\mathbf{p}'-\mathbf{q}'] \, \delta_D^{(1)}[E_1(p)+E_2(q)-E_3(p')-E_4(q')] \\ &\times [f_3(\mathbf{p}')f_4(\mathbf{q}')(1\pm f_3)(1\pm f_4) - f_1(\mathbf{p})f_2(\mathbf{q})(1\pm f_1)(1\pm f_2)] \end{aligned} \tag{6.27}$$

This result is completely general for any 2-particle interaction of the type (6.24) and holds even in the case where some types correspond to the same particle. The microphysics of the interaction is captured by the scattering amplitude $\mathcal{M}(\mathbf{p},\mathbf{q},\mathbf{p}',\mathbf{q}')$.

6.2 The Boltzmann Equation for Photons

The Boltzmann equation to solve is (6.1) for photons, where the left hand side was calculated in (6.22). To go further, one must specify the form of the distribution function, which should reduce to the Bose-Einstein form to zero-th order. To this end, we introduce the perturbation variable in the temperature of photons by writing:

$$T(\mathbf{x},\mathbf{p},t) = T(t) + \delta T(\mathbf{x},\mathbf{p},t) \equiv T(t)(1+\Theta(\mathbf{x},\mathbf{p},t)) \tag{6.28}$$

where the last equivalence is just due to common nomenclature, and the temperature $T(t)$ corresponds to the backround FRW Universe, and as such depends only on time and satisfies $T \propto a^{-1}$ (which is the zero-th order Boltzmann equation, see below). The total distribution function for photons then reads:

$$f(\mathbf{x},p,\hat{\mathbf{p}},t) = \left[\exp\left(\frac{p}{T(t)(1+\Theta(\mathbf{x},\hat{\mathbf{p}},t))}\right) - 1\right]^{-1} \tag{6.29}$$

The perturbation variable Θ then allows for inhomogeneities in the photon distribution (it depends on $\mathbf{x}$) as well as anisotropies (it depends on $\hat{\mathbf{p}}$). Note that the assumption in (6.29) is that the perturbations only depend on the direction of the momentum, but not the magnitude. This will be verified later, and follows directly from the fact that the magnitude of the photon momentum is left unchanged by Compton scattering, which is the dominant form of interaction.
The next step is to expand f everywhere in (6.22) in the parameter Θ to first order. The expansion reads:

$$\begin{aligned} f(\mathbf{x},\mathbf{p},t) &\simeq \frac{1}{e^{p/T(t)}-1} + \left(\frac{\partial}{\partial T}\left[\exp\left(\frac{p}{T(t)}-1\right)^{-1}\right]^{-1}\right) T(t)\Theta(\mathbf{x},\hat{\mathbf{p}},t)) \\ &= f^{(0)}(p,t) - p\frac{\partial f^{(0)}}{\partial p}\Theta(\mathbf{x},\hat{\mathbf{p}},t)) \end{aligned} \tag{6.30}$$

where the zero-th order distribution function is defined as the Bose-Einstein distribution:

$$f^{(0)}(p,t) = \frac{1}{e^{p/T(t)} - 1} \tag{6.31}$$

and in the last line of (6.30) the following identity was used $T\partial f^{(0)}/\partial T = -p\partial f^{(0)}/\partial p$. The left hand side of the Boltzmann equation then becomes an expression for the perturbation in the temperature, where the zero-th order distribution $f^{(0)}$ satisfies:

$$\left(\frac{df}{dt}\right)_{\text{zero-th order}} = \frac{\partial f^{(0)}}{\partial t} - Hp\frac{\partial f^{(0)}}{\partial p} = 0 \tag{6.32}$$

The important point in (6.32) is that the collision term is set to zero, that is the statement that the collision terms will be of first order, i.e. they will be proportional to Θ and other perturbations variables, either in the metric or in the other energy-momentum components. This is a non-trivial statement, but a moment's thought reveals that this is only the condition under which one expects $f^{(0)}$ to be the Bose-Einstein distribution in the first place, that is $f^{(0)}$ is set precisely by the requirement that the collision term vanish. In other words, the collision term includes rates of forward and the corresponding inverse reactions, and if the distribution functions are set to their equilibrium values, the rates cancel. The equation (6.32) is the statement that $T(t) \propto a^{-1}$ in an homogeneous Universe, which is the usual redshift relation. The first order equation instead reads:

$$\begin{aligned}\left(\frac{df}{dt}\right)_{\text{first order}} = &-p\frac{\partial}{\partial t}\left(\frac{\partial f^{(0)}}{\partial p}\Theta\right) - p\frac{\hat{p}^i}{a}\frac{\partial\Theta}{\partial x^i}\frac{\partial f^{(0)}}{\partial p} + H\Theta p\frac{\partial}{\partial p}\left(p\frac{\partial f^{(0)}}{\partial p}\right)\\ &- p\frac{\partial f^{(0)}}{\partial p}\left(\dot{\Phi} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right)\end{aligned} \tag{6.33}$$

Consider the first term on the right hand side:

$$\begin{aligned}-p\frac{\partial}{\partial t}\left(\frac{\partial f^{(0)}}{\partial p}\Theta\right) &= -p\frac{\partial f^{(0)}}{\partial p}\frac{\partial\Theta}{\partial t} - p\frac{\partial T}{\partial t}\frac{\partial^2 f^{(0)}}{\partial T\partial p}\\ &= -p\frac{\partial f^{(0)}}{\partial p}\frac{\partial\Theta}{\partial t} + p\Theta\frac{dT/dt}{T}\frac{\partial}{\partial p}\left(p\frac{\partial f^{(0)}}{\partial p}\right)\end{aligned} \tag{6.34}$$

The second term on the second line cancels with the third term on the right hand side of (6.33), so the full left-hand side of the Boltzmann equation for photons reads:

$$\left(\frac{df}{dt}\right)_{\text{first order}} = -p\frac{\partial f^{(0)}}{\partial p}\left(\dot{\Theta} + \frac{\hat{p}^i}{a}\frac{\partial\Theta}{\partial x^i} + \dot{\Phi} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right) \tag{6.35}$$

The right hand side of the Boltzmann equations regards interactions between photons and other species. The main scattering process of interest in this case is Compton scattering:

$$e^-(\mathbf{q}) + \gamma(\mathbf{p}) \longleftrightarrow e^-(\mathbf{q}') + \gamma(\mathbf{p}') \tag{6.36}$$

From (6.27) the collision term then is:

$$\begin{aligned} C[f(\mathbf{p})] =& \frac{1}{2E(p)} \int \frac{d^3q}{(2\pi)^3 2E_e(q)} \int \frac{d^3q'}{(2\pi)^3 2E_e(q')} \int \frac{d^3p'}{(2\pi)^3 2E(p')} \sum_{3 \text{ spins}} |\mathcal{M}|^2 \\ & \times (2\pi)^4 \, \delta_D^{(3)}[\mathbf{p} + \mathbf{q} - \mathbf{p}' - \mathbf{q}'] \, \delta_D^{(1)}[E(p) + E_e(q) - E(p') - E_e(q')] \\ & \times [f_e(\mathbf{q}') f(\mathbf{p}') - f_e(\mathbf{q}) f(\mathbf{p})] \end{aligned} \tag{6.37}$$

The sum in the scattering amplitude is over the final spin states of the outgoing electron and the photon, and the electron with which the photon with momentum $\mathbf{p}$ scatters. Note that Pauli blocking and stimulated emission were neglected, which would add a term $1 - f_e$ for the electrons and a term $1 + f$ for the photons. Pauli blocking is never important after electron-positron annihilation, since the occupation numbers f_e are generally small. On the other hand, the stimulated emission factor drops out, as we will see briefly. Note also the notation: we have used f to indicate the distribution function for photons, and f_e the one for electrons.
The photon energies are obviously $E(p) = p$ and $E(p') = p'$, whereas for electrons we assume the non-relativistic limit, which is a valid assumptions at the time of recombination, where typical kinetic energies of electrons, of order $T_{\rm rec} \simeq 10^{-1}$ eV are much smaller than the electron mass $m_e \simeq 1$ MeV. Then:

$$\begin{aligned} E(p) = p &\sim T \\ E_e(q) - m_e = \frac{q^2}{2m_e} &\sim T \quad \Longrightarrow \quad q \sim T\sqrt{\frac{2m_e}{T}} \end{aligned} \tag{6.38}$$

It is clear that electron momenta are much larger than photon momenta, since, as argued above, $m_e/T \gg 1$. Now, one can use the three dimensional delta function in (6.37) to solve the $\mathbf{q}'$ integral:

$$\begin{aligned} C[f(\mathbf{p})] =& \frac{\pi}{2m_e p} \int \frac{d^3q}{(2\pi)^3 2m_e} \int \frac{d^3p'}{(2\pi)^3 2p'} \delta_D^{(1)}[p + E_e(q) - p' - E_e(|\mathbf{q} + \mathbf{p} - \mathbf{p}'|)] \\ & \times \sum_{3 \text{ spins}} |\mathcal{M}|^2 [f_e(\mathbf{q} + \mathbf{p} - \mathbf{p}') f(\mathbf{p}') - f_e(\mathbf{q}) f(\mathbf{p})] \end{aligned} \tag{6.39}$$

The next approximation comes from analysing the dynamics of non-relativistic Compton scattering, the most important feature of which is the little to no energy transfer. In particular:

$$p - p' = E_e(q) - E_e(|\mathbf{q} + \mathbf{p} - \mathbf{p}'|) = \frac{q^2}{2m_e} - \frac{|\mathbf{q} + \mathbf{p} - \mathbf{p}'|^2}{2m_e} \simeq \frac{|\mathbf{p}' - \mathbf{p}| \cdot \mathbf{q}}{m_e} \tag{6.40}$$

where the last equality holds because, as argued above, q is much larger than p and p'. Now, since p and p' are of the same order (the photons are thermal), the right hand side (6.40) is at most $2pq/m_e$, so the fractional change in photon energy is at most $(p - p')/p \leq 2q/m_e \sim 2\sqrt{2T/m_e} \ll 1$. Thus, the non-relativistic Compton scattering is elastic and $p \sim p'$, which justifies the assumption that Θ only depends on $\hat{\mathbf{p}}$ but not on p. Finally, it is now possible to expand the final electron kinetic energy $|\mathbf{q} + \mathbf{p} - \mathbf{p}'|^2/2m_e$ around the value $q^2/2m_e$, so the one-dimensional delta

function becomes:

$$\delta_D^{(1)}[p + E_e(q) - p' - E_e(|\mathbf{q} + \mathbf{p} - \mathbf{p}'|)] \simeq \delta_D^{(1)}(p - p') + \frac{(\mathbf{p} - \mathbf{p}') \cdot \mathbf{q}}{m_e} \frac{\partial}{\partial p'} \delta_D^{(1)}(p - p') \tag{6.41}$$

The derivative of the delta function will be handled via integration by parts as usual. Using this expansion and the fact that $f_e(\mathbf{q} + \mathbf{p} - \mathbf{p}') \simeq f_e(\mathbf{q})$, the collision term becomes:

$$\begin{aligned} C[f(\mathbf{p})] = & \frac{\pi}{8 m_e^2 p} \int \frac{d^3q}{(2\pi)^3} f_e(\mathbf{q}) \int \frac{d^3p'}{(2\pi)^3 p'} \sum_{3 \text{ spins}} |\mathcal{M}|^2 \\ & \times \left[\delta_D^{(1)}(p - p') + \frac{(\mathbf{p} - \mathbf{p}') \cdot \mathbf{q}}{m_e} \frac{\partial \delta_D^{(1)}(p - p')}{\partial p'} \right] \end{aligned} \tag{6.42}$$

It is obvious now why the stimulated emission terms were neglected at first. Including them would result in changing in the final factor to $[f(\mathbf{p}')(1 + f(\mathbf{p})) - f(\mathbf{p})(1 + f(\mathbf{p}'))]$, where the additional terms cancel out. The final step is to lay out the expression for the scattering amplitude. The low-energy limit of Compton scattering results in:

$$\frac{1}{2} \sum_{4 \text{ spins}} |\mathcal{M}|^2 = 24 \pi \sigma_T m_e^2 \left(1 + [\mathbf{p} \cdot \mathbf{p}']^2\right) \tag{6.43}$$

where σ_T is the Thomson cross-section. Summing over the polarization states of the photon with momentum $\mathbf{p}$ as well (the prefactor of $1/2$), means that we can use this expression for $\sum_{3 \text{ spins}} |\mathcal{M}|^2$. A further simplification consists in averaging over the angle in the parenthesis:

$$\sum_{3 \text{ spins}} |\mathcal{M}|^2 = 32 \pi \sigma_T m_e^2 \tag{6.44}$$

Since the angular dependence would result in a numerically subdominant factor in the final collision term, it is reasonable to neglect it. Note that the average over spin states of the ingoing and outgoing photons ignores polarization effects in the radiation field. In reality, the amplitude for the Compton scattering has a polarization dependence, which leads to polarization in the CMB. This effect contains extremely valuable information, especially with regards to "B-mode" polarization, which is induced by tensor perturbations (gravitational waves) in the early Universe. We will neglect polarization effects for the time being.

Once established that the scattering amplitude is momenta-independent, the terms in the brackets in (6.42) can be safely multiplied an integrated over q. The integral then simply gives a factor of $n_e/2$ for the terms independent of q. Terms which contain a factor of $\mathbf{q}/m_e$ instead result in $n_e \mathbf{u}_b/2$, where u_b is the bulk velocity of the electrons (the subscript b stands for baryons, but protons are closely coupled to

electrons anyway. See below). Therefore:

$$\begin{aligned} C[f(\mathbf{p})] &= \frac{2\pi^2 n_e \sigma_T}{p} \int \frac{d^3p'}{(2\pi)^3 p'} \left[\delta_D^{(1)}(p-p') + (\mathbf{p}-\mathbf{p}') \cdot \mathbf{u}_b \frac{\partial \delta_D^{(1)}(p-p')}{\partial p'} \right] \\ &\times \left[f^{(0)}(p') - f^{(0)}(p) - p' \frac{\partial f^{(0)}}{\partial p'} \Theta(\hat{\mathbf{p}}') + p \frac{\partial f^{(0)}}{\partial p} \Theta(\hat{\mathbf{p}}) \right] \\ &= \frac{n_e \sigma_T}{4\pi p} \int_0^\infty dp' \; p' \int d\Omega' \; [\delta_D^{(1)}(p-p') \left(-p \frac{\partial f^{(0)}}{\partial p'} \Theta(\hat{\mathbf{p}}') + p \frac{\partial f^{(0)}}{\partial p} \Theta(\hat{\mathbf{p}}) \right) \\ &+ (\mathbf{p}-\mathbf{p}') \cdot \mathbf{u}_b \frac{\partial \delta_D^{(1)}(p-p')}{\partial p'} (f^{(0)}(p') - f^{(0)}(p)] \end{aligned} \tag{6.45}$$

where Ω' is the solid angle spanned by the momentum unit vector $\hat{\mathbf{p}}'$. Again, notice that the $(\mathbf{x}, t)$ dependence has been omitted in the right hand side, since all collisions are local. There are now only two terms which depend on $\hat{\mathbf{p}}'$ and are therefore affected by integration. First, the perturbation term $\Theta(\hat{\mathbf{p}}')$ is handled by introducing the monopole of the temperature distribution:

$$\Theta_0(\mathbf{x}, t) \equiv \frac{1}{4\pi} \int d\Omega' \; \Theta(\hat{\mathbf{p}}', \mathbf{x}, t) \tag{6.46}$$

The monopole corresponds to the angle averaged flux of photons at any given point in space $\mathbf{x}$ and at time t. This can be generalized to a sequence of multipole moments Θ_l, which are integrals of the perturbation Θ, weighted by different functions of the directions $\hat{p}$. The second term that depends on $\hat{\mathbf{p}}'$ is the factor $\hat{\mathbf{p}}' \cdot \mathbf{u}_b$, which integrates to zero since $\mathbf{u}_b$ is independent of both $\hat{\mathbf{p}}'$ and $\mathbf{p}$. The integration over Ω' then gives:

$$\begin{aligned} C[f(\mathbf{p})] = \frac{n_e \sigma_T}{p} \int_0^\infty dp' p' \; [\delta_D^{(1)}(p-p') \left(-p \frac{\partial f^{(0)}}{\partial p'} \Theta_0 + p \frac{\partial f^{(0)}}{\partial p} \Theta(\hat{\mathbf{p}}) \right) \\ + \mathbf{p} \cdot \mathbf{u}_b \frac{\partial \delta_D^{(1)}(p-p')}{\partial p'} (f^{(0)}(p') - f^{(0)}(p))] \end{aligned} \tag{6.47}$$

Now the final integral over p' can be carried out, the first line by using trivially the delta function, and the second line with integration by parts:

$$C[f(\mathbf{p})] = -p \frac{\partial f^{(0)}}{\partial p} n_e \sigma_T [\Theta_0 - \Theta(\hat{\mathbf{p}}) + \hat{\mathbf{p}} \cdot \mathbf{u}_b] \tag{6.48}$$

From this equation the effects of efficient Compton scattering should be clear. In the absence of an electron bulk velocity, i.e. $\mathbf{u}_b = 0$, the collision term serves to drive Θ to Θ_0. This should be intuitive, since if the scattering is very effective, then the photon mean free path should be very small, meaning that a photon at any given point in space will have scattered off electrons that were very nearby. These electrons had a temperature very similar to the local one, so the global temperature distribution is almost completely described by its monopole. A slight caveat is that a bulk electron velocity naturally introduces a dipole in the photon distribution Θ, so that techincally it is described by its monopole Θ_0 and its dipole Θ_1, that is to say that the

photons behave exactly like a fluid. Indeed, the monopole is related to the density field, whereas the dipole is related to the velocity field. Tight coupling between photons and baryons then creates a situation wherein the photons and baryons behave as a single fluid. This will not be true after decoupling, when the photons are free to stream in the Universe. This situation will therefore need to include all multipole moments of the photon perturbation distribution, but will still be handled by the Boltzmann approach.
The left and right hand sides of the Boltzmann equation for photons, (6.35) and (6.48), can be collected into the following equation:

$$\dot{\Theta} + \frac{\hat{p}^i}{a}\frac{\partial \Theta}{\partial x^i} + \dot{\Phi} + \frac{\hat{p}^i}{a}\frac{\partial \Psi}{\partial x^i} = n_e \sigma_T [\Theta_0 - \Theta(\hat{\mathbf{p}}) + \hat{\mathbf{p}} \cdot \mathbf{u}_b] \tag{6.49}$$

It is convenient to work in conformal time η, so that the equation becomes:

$$\Theta' + \hat{p}^i \frac{\partial \Theta}{\partial x^i} + \Phi' + \hat{p}^i \frac{\partial \Psi}{\partial x^i} = n_e \sigma_T a [\Theta_0 - \Theta(\hat{\mathbf{p}}) + \hat{\mathbf{p}} \cdot \mathbf{u}_b] \tag{6.50}$$

where the primes denote derivatives with respect to η (while dots denote derivatives with respect to comoving time t). Equation (6.50) is a partial differential equation in the perturbation variable Θ, which is coupled to the variables $\mathbf{u}_b$, Φ and Ψ, which satisfy differential equations themselves that will be derived below. As usual, it is convenient to switch to Fourier space when dealing with differential equations. In the case at hand, that is the Boltzmann/Einstein equations at linear order, an additional benefit is that the modes k do not mix in the equations.
Instead of working with the the photon momenta $\mathbf{p}$, one uses the angle between the wavenumber $\mathbf{k}$ and the photon direction $\hat{\mathbf{p}}$, defined as:

$$\mu \equiv \frac{\mathbf{k} \cdot \hat{\mathbf{p}}}{k} \tag{6.51}$$

In this way, since, by definition, the wavevector $\mathbf{k}$ points in the direction in which the temperature is changing, $\Theta(k, \mu = 1)$ describes photons traveling in the direction of the temperature gradient, whereas $\Theta(k, \mu = 0)$ describes photons traveling perpendicularly to it, i.e. along a path along which the temperature does not change. Furthermore, in Cosmology velocities are generally longitudinal, that is they point in the direction of $\mathbf{k}$:

$$\mathbf{u}_b(\mathbf{k}, \eta) = \frac{\mathbf{k}}{k} u_b(\mathbf{k}, \eta) \tag{6.52}$$

so $\mathbf{u}_b \cdot \hat{\mathbf{p}} = u_b \mu$. Next, the optical depth is defined as:

$$\tau(\eta) \equiv \int_{\eta}^{\eta_0} d\eta' \, n_e \sigma_T a \tag{6.53}$$

The limits of integrations are set such that:

$$\tau' = -n_e \sigma_T a \tag{6.54}$$

With the new definitions (6.51) and (6.53), the photon Boltzmann equation becomes finally:

$$\Theta' + ik\mu\Theta + \Phi' + ik\mu\Psi = -\tau'[\Theta_0 - \Theta + \mu u_b] \tag{6.55}$$

6.3 The Boltzmann Equation for Cold Dark Matter

The next component we analyse with the Boltzmann equation formalism is dark matter. By definition, it is collisionless and consists of massive particles, so the correct equation to use is the one derived in (6.23):

$$\frac{\partial f_c}{\partial t} + \frac{p}{E}\frac{\hat{p}^i}{a}\frac{\partial f_c}{\partial x^i} - \left[H + \dot{\Phi} + \frac{E}{ap}\hat{p}^i\partial_i\Psi\right] p\frac{\partial f_c}{\partial p} = 0 \tag{6.56}$$

When treating dark matter perturbations, one should keep in mind that dark matter is non-relativistic, which simplifies the equations greatly. In practice, while for the photon Boltzmann equation the knowledge of the zero-th order distribution was instrumental, here the fact that particles are non-relativistic is all that's needed, since increasingly higher-order powers of p will be neglected. Only terms with $(p/m)^0$ and $(p/m)^1$ will be retained, meaning that the bulk motion will be included, but not the velocity dispersion. This is tantamount to treating CDM as an effective fluid (see below).
Without assuming a form for f_c, one takes moments of the master equation (6.56). The zero-th order moment is obtained by multiplying both sides by the phase space volume $d^3p/(2\pi)^3$ and integrating:

$$\begin{aligned}\frac{\partial}{\partial t}\int\frac{d^3p}{(2\pi)^3}\,f_c + \frac{1}{a}\frac{\partial}{\partial x^i}\int\frac{d^3p}{(2\pi)^3}\,f_c\frac{p\hat{p}^i}{E(p)} - [H+\dot{\Phi}]\int\frac{d^3p}{(2\pi)^3}\,p\frac{\partial f_c}{\partial p} \\ - \frac{1}{a}\frac{\partial\Psi}{\partial x^i}\int\frac{d^3p}{(2\pi)^3}\,\frac{\partial f_c}{\partial p}E(p)\hat{p}^i = 0\end{aligned} \tag{6.57}$$

Integration by part shows that the last term vanishes, while for the first two terms one uses simply definitions:

$$\begin{aligned}n_c &= \int\frac{d^3p}{(2\pi)^3}\,f_c \\ u_c &= \frac{1}{n_c}n_c = \int\frac{d^3p}{(2\pi)^3}\,f_c\frac{p\hat{p}^i}{E(p)}\end{aligned} \tag{6.58}$$

where n_c is the number density and u_c is the bulk velocity. One should be careful of the distinction between fluid bulk velocities $\mathbf{u}$ and velocities of single particles. The two are generally different: while single velocities may be large, the bulk motion that is produced as an average may be small. The third term in (6.57) can be integrated by parts:

$$\int\frac{d^3p}{(2\pi)^3}\,p\frac{\partial f_c}{\partial p} = \frac{1}{(2\pi)^3}\int_0^\infty dp\,p^3\frac{\partial}{\partial p}\int d\Omega\,f_c = -3\frac{1}{(2\pi)^3}\int dp\,p^2\int d\Omega\,f_c = -3n_c \tag{6.59}$$

The zero-th moment of the Boltzmann equation for matter is therefore the perturbed continuity equation:

$$\frac{\partial n_c}{\partial t} + \frac{1}{a}\frac{\partial(n_c u_c^i)}{\partial x^i} + 3[H+\dot{\Phi}]n_c = 0 \tag{6.60}$$

To extract a first order equation from (6.60), one notices that u_c is already a first order term, and n_c is instead given by:

$$n_c(\mathbf{x},t) = \bar{n}_c(t)(1+\delta_c(\mathbf{x},t)) \tag{6.61}$$

where $\bar{n}_c(t)$ is the averaged out number density and obeys the equation:

$$\frac{\partial \bar{n}_c}{\partial t} + 3H\bar{n}_c(t) = 0 \tag{6.62}$$

which is the same equation as for ρ_c. The solution implies the scaling $\bar{n}_c(t) \propto a^{-3}$, as is expected in a FRW Universe. The term δ_c is instead a new perturbation variable that one should in principle solve for, and is called the "density contrast", since $\delta n_c/\bar{n}_c = \delta\rho_c/\bar{\rho}_c$. Plugging (6.61) into (6.60) and keeping only first order terms yields:

$$\frac{\partial \delta c}{\partial t} + \frac{1}{a}\frac{\partial u_c^i}{\partial x^i} + 3\dot{\Phi} = 0 \tag{6.63}$$

Since there are two variables in this equation, we need to derive another equation to close the system. To get it, one takes the first moment of (6.23), taking care to neglect the terms of order $(p/m)^2$. Multiplying by $d^3p/(2\pi)^3\ p\hat{p}^j/E$ both sides of (6.23) and integrating, one has:

$$\frac{\partial}{\partial t}\int \frac{d^3p}{(2\pi)^3}\, f_c \frac{p\hat{p}^j}{E} - [H+\dot{\Phi}]\int \frac{d^3p}{(2\pi)^3}\frac{\partial f_c}{\partial p}\frac{p^2\hat{p}^j}{E} - \frac{1}{a}\frac{\partial \Psi}{\partial x^i}\int \frac{d^3p}{(2\pi)^3}\frac{\partial f_c}{\partial p} p\hat{p}^i\hat{p}^j = 0 \tag{6.64}$$

The first term in (6.64) is just the time derivative of $n_c u_c^i$. The second term instead is handled with integration by parts:

$$\begin{aligned}\int \frac{d^3p}{(2\pi)^3}\frac{\partial f_c}{\partial p}\frac{p^2\hat{p}^j}{E} &= \int \frac{d\Omega}{(2\pi)^3}\,\hat{p}^j \int_0^\infty dp\ \frac{p^4}{E}\frac{\partial f_c}{\partial p} \\ &= -\int \frac{d\Omega}{(2\pi)^3}\,\hat{p}^j \int_0^\infty dp\ f_c\left(\frac{4p^3}{E} - \frac{p^5}{E^3}\right)\end{aligned} \tag{6.65}$$

The first term in the second line $\propto -4p^3/E$ yields $-4n_c u_c^i$ upon integration, while the term involving p^5/E^3 can be neglected. The same steps carry through for the third term in (6.64), but one has to use the additional integral:

$$\int d\Omega\ \hat{p}^i\hat{p}^j = \delta^{ij}\frac{4\pi}{3} \tag{6.66}$$

The first moment of the CDM Boltzmann equation then becomes:

$$\frac{\partial(n_c u_c^j)}{\partial t} + 4Hn_c u_c^j + \frac{n_c}{a}\frac{\partial \Psi}{\partial x^j} = 0 \tag{6.67}$$

This equation has no zero-th order part, since the velocity is already a first order perturbation, as above. Therefore it is only sufficient to make the substitution $n_c \to \bar{n}_c$ everywhere. Using the time dependence $\bar{n}_c \propto a^{-3}$, one then finds:

$$\frac{\partial u_c^j}{\partial t} + Hu_c^j + \frac{1}{a}\frac{\partial \Psi}{\partial x^j} = 0 \tag{6.68}$$

This is the perturbed first order Euler equation for a fluid. These two equations, (6.63) and (6.68), govern the evolution of the density and velocity of cold dark matter, which in turn depend on the metric perturbations Φ and Ψ. An interesting fact of these two equations is generic to this process of extracting the fluid equations from the Boltzmann equations: the integrated Boltzmann equation for the $l-$th moment depends on the moment of order $l+1$, that is, for example, the equation for the density contrast (the zero-th moment of the distribution function) depends on the bulk velocity (the first moment of the distribution function). The process of integration, as carried out in this section, leads to an infinite hierarchy of equations for the moments $l = 0, 1, 2...$. This is the case of photons, for which there are an infinite amount of equations governing the multipoles Θ_l. In the case of CDM, however, the hierarchy is closed due to the non-relativistic nature of dark matter, which sets to zero all the moments $l > 1$. This, again, corresponds to treating dark matter as a fluid, and indeed the two equations (6.63) and (6.68) can be derived with the conservation equations $\nabla_\mu T^{\mu\nu} = 0$, where $T^{\mu\nu}$ contains the appropriate perturbations for a perfect fluid. For massive particles with larger velocities, such as neutrinos, the hierarchy cannot be closed, and one needs to keep track of higher moments.
As a final step, we switch to Fourier space and conformal time, such that the continuity and Euler equations read:

$$\begin{aligned} &\delta_c' + iku_c + 3\Phi' = 0 \\ &u_c' + \frac{a'}{a}u_c + ik\Psi = 0 \end{aligned} \tag{6.69}$$

6.4 The Boltzmann Equation for Baryons

The small percentage of matter that is not dark is made of electrons, protons, neutrons and heavier nuclei, like helium and lithium. They are collectively denoted baryons because the rest masses that dominate are those of protons and neutrons out of all the particles cited. Electrons and protons are tightly coupled by Coulomb scattering ($e + p \leftrightarrow e + p$), and remain coupled even after recombination, since they combine to form atoms. The Coulomb scattering rate, before recombination, is much larger than the expansion rate and therefore the electron and proton density contrasts are the same:

$$\frac{\rho_e - \bar{\rho}_e}{\bar{\rho}_e} = \frac{\rho_b - \bar{\rho}_b}{\bar{\rho}_b} \equiv \delta_b \tag{6.70}$$

Similarly, the velocities of the two species are forced to a common value too:

$$\mathbf{u}_e = \mathbf{u}_p \equiv \mathbf{u}_b \tag{6.71}$$

Because of this tight coupling and correspondingly small mean free paths, and the fact that they are non-relativistic since $T \ll m_e$, electrons and nuclei can be treated as a non-relativistic fluid and the Boltzmann hierarchy closes like before, that is we will only be concerned with the first and second moments of the Boltzmann equation. This time, however, a collision term in the form of Coulomb scattering will need to be introduced.
The procedure of extracting the first and second moments of the Boltzmann equation

are the same as for dark matter, in particular the zero-th moment for baryons yields:

$$\delta_b' + ik\mu u_b + 3\Phi' = 0 \tag{6.72}$$

Note that, despite collisions, the right hand side was put to zero. This is because collisions preserve the number of electrons and protons, which is true for Compton and Coulomb scattering, but not for pair production or annihilation reactions; the latter are, however, irrelevant around and after recombination. The continuity equation with vanishing collision term captures this number conservation.
The second equation, the one for $\mathbf{u}_b$, is derived by taking the second moments of the Boltzmann equations for electrons and protons and baryons and summing them all together. Also, this time we multiply by $1/E$ and not $\mathbf{p}/E$ before integrating, so that the final equation will look like the left hand side of (6.67), except it will be multiplied by m_p, since it is the dominant mass among baryons/electrons. Therefore, one gets:

$$m_p \frac{\partial (n_b u_b^j)}{\partial t} + 4H m_p n_b u_b^j + \frac{m_p n_b}{a}\frac{\partial \Psi}{\partial x^j} = F_{e\gamma}^j(\mathbf{x}, t) \tag{6.73}$$

The collision term is not zero, since Compton scattering does not preserve the momentum of particles. This momentum transfer between electrons and photons is then captured by the term $F_{e\gamma}^j$, and the electrons in turn transfer their momenta immediately to the nuclei. Technically, photons also scatter off nuclei, but the cross section is suppressed by a factor $(m_e/m_p)^2 \sim 10^{-6}$ compared to that of electron-photon scattering. Dividing both sides by $\rho_b = m_b \bar{n}_b$ we get:

$$\frac{\partial u_b}{\partial t} + H u_b^j + \frac{1}{a}\frac{\partial \Psi}{\partial x^j} = \frac{1}{\rho_b} F_{e\gamma}^j(\mathbf{x}, t) \tag{6.74}$$

The last step in the derivation is to evaluate the collision term $F_{e\gamma}^j$, for which a shortcut is conveniently used. Since $F_{e\gamma}^j$ describes momentum transfer between electrons and photons, and total momentum is conserved in each scattering event, this force term has to be precisely equal and opposite to the force term appearing in the photon Boltzmann equation (6.48). To get the baryon collision term, we have to take the first moment of it. First, translating (6.48) to Fourier space and anticipating $F_{e\gamma}^j$ to be aligned with the wavevector $\mathbf{k}$, we multiply the equation by $\hat{k}^j$ before integrating over phase space. In addition, since the momentum $n_e u_e^i$ of electrons technically counts both spin states (it is twice the first moment of the distribution function), the collision term must be multiplied by two. Together with the minus sign from momentum conservation, we then multiply (6.48) by $-2p\mu$ and integrate over $\mathbf{p}$, the photon momentum:

$$\begin{aligned} \frac{1}{\rho_b}\hat{k}^i F_{e\gamma}^i &= -\frac{2n_e\sigma_T}{\rho_b}\int \frac{d^3p}{(2\pi)^3}\, p\mu \left[-p\frac{\partial f^{(0)}}{\partial p}\right][\Theta_0 - \Theta(\mu) + \mu u_b] \\ &= \frac{2n_e\sigma_T}{\rho_b}\int_0^\infty \frac{dp}{2\pi^2}\, p^4 \frac{\partial f^{(0)}}{\partial p} \int_{-1}^{+1}\frac{d\mu}{2}\,\mu[\Theta_0 - \Theta(\mu) + \mu u_b] \end{aligned} \tag{6.75}$$

The integral over p can be done by parts and yields $-2\rho_\gamma$, since the background radiation energy density is twice the integral of $pf^{(0)}(p)$. The integration over μ, instead vanishes for the first term, and gives $u_b/3$ for the third term. The second

term, on the other hand gives the first moment of the temperature perturbation Θ, which is defined as:

$$\Theta_1(k,\eta) \equiv i\int_{-1}^{+1}\frac{d\mu}{2}\Theta(\mu,k,\eta) \tag{6.76}$$

Finally, the collision term becomes:

$$\frac{1}{\rho_b}\hat{k}^i F^i_{e\gamma}(\mathbf{x},t) = -n_e\sigma_T\frac{4\rho_\gamma}{\rho_b}\left[i\Theta_1+\frac{1}{3}u_b\right] \tag{6.77}$$

To understand the physical significance of the appearance of the first moment Θ_1 in the Euler equation, remember that $F^i_{e\gamma}$ accounts for momentum transfer between photons and electrons, which is clearly zero in an isotropic temperature field. On the other hand, if there is a non-zero dipole term, more energetic photons come from one direction as opposed to the opposite direction. Electrons moving against this flow will then experience a head wind, a force pointing in the opposite direction, called the "Compton drag".
The full second baryon Boltzmann equation then is:

$$u_b' + \frac{a'}{a}u_b + ik\Psi = \tau'\frac{4\rho_\gamma}{3\rho_b}[3i\Theta_1+u_b] \tag{6.78}$$

The reason why the baryon energy density ρ_b appears in this equation despite photons scattering primarily off electrons can be understood as follows. Electrons are difficult to move because of their tight coupling to baryons via Coulomb scattering, indeed if $\rho_b \to \infty$, Compton scattering would not change electron velocity at all. Finally, the results were derived by assuming hydrogen to be completely ionized, that is $n_e = n_p = n_b$. However, the result turns out to be valid even in the presence of an appreciable amount of neutral hydrogen atoms, i.e. $n_e < n_b$.

6.5 The Boltzmann Equation for Neutrinos

The last species of particles we analyse is neutrinos, with distribution function $f_\nu(\mathbf{x},\mathbf{p},t)$. At zero-th order, neutrinos follow an equilibrium distribution function with temperature $T_\nu(a)$, so the same treatment of photons can be applied. Then, a perturbation in the neutrino temperature distribution is inserted:

$$T_\nu(\mathbf{x},\mathbf{p},t) = T_\nu(t)(1+\mathcal{N}(\mathbf{x},\mathbf{p},t)) \tag{6.79}$$

so that f_ν takes the form:

$$f_\nu(\mathbf{x},\mathbf{p},t) = \left[\exp\left(\frac{p}{T_\nu(t)(1+\mathcal{N}(\mathbf{x},\mathbf{p},t))}\right)+1\right]^{-1} \tag{6.80}$$

which can be expanded to first order as:

$$f_\nu(\mathbf{x},\mathbf{p},t) \simeq \left[1-\mathcal{N}(\mathbf{x},\mathbf{p},t)p\frac{d}{dp}\right]f_\nu^{(0)}(p) \tag{6.81}$$

where, again, the zero-th order distribution function is of the Bose-Einstein form $f^{(0)} = [e^{p/T_\nu(a)} + 1]^{-1}$. During the epoch of interest, that is after neutrino decoupling, any non-gravitational interaction between neutrinos and other species can be neglected, so the correct Boltzmann equation is (6.23), which we recall here below:

$$\frac{\partial f_\nu}{\partial t} + \frac{p}{E_\nu(p)}\frac{\hat{p}^i}{a}\frac{\partial f_\nu}{\partial x^i} - \left[H + \dot{\Phi} + \frac{E_\nu(p)}{ap}\hat{p}^i\partial_i\Psi\right] p\frac{\partial f_\nu}{\partial p} = 0 \tag{6.82}$$

Plugging the expansion of f_ν to first order into (6.82) and keeping only first order terms yields an equation for $\mathcal{N}$:

$$\frac{\partial \mathcal{N}}{\partial t} + \frac{p}{E_\nu(p)}\frac{\hat{p}^i}{a}\frac{\partial \mathcal{N}}{\partial x^i} - Hp\frac{\partial \mathcal{N}}{\partial p} + \dot{\Phi} + \frac{E_\nu(p)}{ap}\hat{p}^i\partial_i\Psi = 0 \tag{6.83}$$

Converting now to Fourier space and using conformal time:

$$\mathcal{N}'(\mathbf{k}, p, \mu, \eta) + ik\mu\frac{p}{E_\nu(p)}\mathcal{N} - Hp\frac{\partial}{\partial p}\mathcal{N} = -\Phi' - ik\mu\frac{E_\nu(p)}{p}\Psi \tag{6.84}$$

Apart from the missing collision term, this equation differs from that of photons primarily through the factors $p/E_\nu(p)$, which obviously reduce to unity when the neutrinos are ultra relativistic. At late times, when the factor deviates from unity, the first term on the left proportional to $p/E_\nu(p)$ reflects the slow-down in free streaming due to the mass, whereas the last term on the right hand side reflects the fact that slow-moving neutrinos spend more time in potential wells compared to photons, and are therefore more affected by them. Furthermore, one can no longer assume $\mathcal{N}$ to be independent of p, like Θ, because the factor $p/E_\nu(p)$ assumes very different values for neutrinos in different parts of the distribution. For this same reason, one would need different distributions for the three different generations of neutrinos. However, if one is only interested in the behavior of neutrinos only up to recombination, then $p/E_\nu(p) = 1$ and the p dependence can be safely neglected.

Chapter 7

The Einstein Equations in a Perturbed Universe

In the previous chapter, the equations for the matter-energy perturbations were derived while taking into account the effects of gravity by specifying a perturbed metric. To close out the system of equations, we need to find equations that control the behavior of the metric perturbations, that is how the matter-energy content influences space-time. These equations are the perturbed Einstein equations.

7.1 The Gauge Philosophy

In the perturbation theory of General Relativity, one considers a perturbed spacetime that is close to being a simple, symmetric spacetime, the background spacetime. During our discussion, these are the two objects that constantly play with each other:

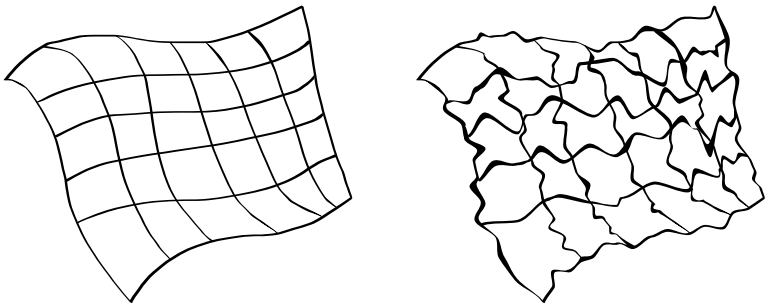

FIGURE 7.1: The two spacetimes at play in perturbation theory: the background spacetime on the left and the total, perturbed one on the right

The coordinate system on the perturbed spacetime can be written with the following metric:

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu} \tag{7.1}$$

where $\bar{g}_{\mu\nu}$ is the metric on the background, unperturbed spacetime and $\delta g_{\mu\nu}$ is just a small perturbation. We require in general, even though it's not strictly necessary, that the first and second derivatives of the perturbation $\delta g_{\mu\nu,\rho}$ and $\delta g_{\mu\nu,\rho\sigma}$ are small. The same idea applies to the Einstein and energy momentum tensors:

$$\begin{aligned} G_{\mu\nu} &= \bar{G}_{\mu\nu} + \delta G_{\mu\nu} \\ T_{\mu\nu} &= \bar{T}_{\mu\nu} + \delta T_{\mu\nu} \end{aligned} \tag{7.2}$$

The background and the whole (perturbed) spacetimes obey the Einstein equations:

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \quad \text{and} \quad \bar{G}_{\mu\nu} = 8\pi G \bar{T}_{\mu\nu} \tag{7.3}$$

From these we can see, by subtracting them, that the perturbations obey the same equations:

$$\delta G_{\mu\nu} = 8\pi G \delta T_{\mu\nu} \tag{7.4}$$

We need to make a few remarks here. First, the above calculations require a pointwise correspondence in the two spacetimes given by the coordinate system $\{x^\alpha\}$, which lets us subtract the two equations, in the following sense: the point $\bar{P}$ in the background and the mapped one P in the perturbed spacetime correspond (i.e. are the same physical point) and have the same coordinate. Given a background spacetime though, there exist many coordinate systems, very close to one another, for which (7.1) holds. The choice among all these coordinate systems is called a gauge choice, and it is of critical importance in perturbative GR.
Secondly, the fact that the spatial averages $\bar{G}_{\mu\nu}$ and $\bar{T}_{\mu\nu}$ satisfy the Einstein equations is only true in linear perturbation theory. In general, we expect this to not be true since GR is a nonlinear theory, so that the perturbations affect the evolution of the mean. This effect is called backreaction and the degree to which it influences the background is still a subject of intense research. In fact, since by the time of structure formation these perturbation grow very large, we expect this backreaction to affect non-trivially the evolution of the background universe.
Now that we've seen this separation of quantities, we have to tackle the imporant problem of gauge transformations since, as we said, there are many possible perturbed systems for a given background spacetime, all close to one another. In this case, a gauge transformation means a coordinate transformation that relates two of these perturbed systems. To develop the formalism, we denote the coordinates of the background by x^α and the ones in the two different perturbed systems as $\hat{x}^\alpha$ and $\tilde{x}^\alpha$; the latter are related then by a gauge transformation by:

$$\tilde{x}^\alpha = \hat{x}^\alpha + \xi^\alpha \tag{7.5}$$

where ξ is a small parameter (we thus drop products of this with itself, and its derivatives).

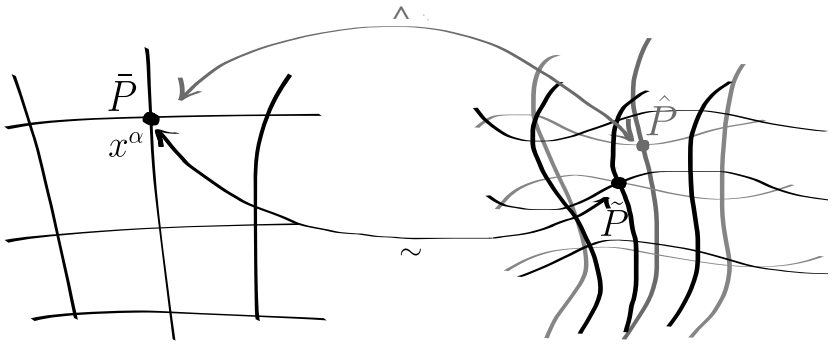

FIGURE 7.2: A gauge transformation. On the left, the same background spacetime is mapped on two different perturbed ones, which are related by a coordinate transformation.

In particular, the difference between $\tilde{\partial}_\beta \xi^\alpha$ and $\hat{\partial}_\beta \xi^\alpha$ is second order small and

thus ignored, so we can actually think of ξ as living in the background spacetime. We illustrate the situation in Figure 7.2. We know that the coordinate system $\{\hat{x}^\alpha\}$ associates point $\bar{P}$ in the background with $\hat{P}$, whereas $\{\tilde{x}^\alpha\}$ associates it to $\tilde{P}$. This is done by:

$$\tilde{x}^\alpha(\tilde{P}) = \hat{x}^\alpha(\hat{P}) = x^\alpha(\bar{P}) \tag{7.6}$$

The gauge transformation relates the coordinates of the same point it the perturbed spacetime, that is:

$$\begin{aligned} \tilde{x}^\alpha(\tilde{P}) &= \hat{x}^\alpha(\tilde{P}) + \xi^\alpha \\ \tilde{x}^\alpha(\hat{P}) &= \hat{x}^\alpha(\hat{P}) + \xi^\alpha \end{aligned} \tag{7.7}$$

where we notice that the difference between $\xi^\alpha(\tilde{P})$ and $\xi^\alpha(\hat{P})$ is second order small, and thus we associate ξ^α to the background:

$$\xi^\alpha = \xi^\alpha(\bar{P}) = \xi^\alpha(x^\beta) \tag{7.8}$$

Putting together equations (7.6) and (7.7) we get the relation between two different points in the same coordinate system:

$$\begin{aligned} \hat{x}^\alpha(\tilde{P}) &= \hat{x}^\alpha(\hat{P}) - \xi^\alpha \\ \tilde{x}^\alpha(\tilde{P}) &= \tilde{x}^\alpha(\hat{P}) - \xi^\alpha \end{aligned} \tag{7.9}$$

We are now ready to understand how various tensors transform under a gauge transformation. We start with the perturbed quantities:

$$\begin{aligned} s &= \bar{s} + \delta s \\ w^\alpha &= \bar{w}^\alpha + \delta w^\alpha \\ B_{\alpha\beta} &= \bar{B}_{\alpha\beta} + \delta B_{\alpha\beta} \end{aligned} \tag{7.10}$$

In the following, we derive how the scalar s changes under a gauge transformation, and just state the results for the other two cases. Let's first consider the full quantity $s = \bar{s} + \delta s$, which lives on the perturbed spacetime. We cannot assign a unique background quantity $\bar{s}$ to a point in the perturbed spacetime, because different gauges may have the same values for for $\bar{s}$. This implies that there is no unique perturbation δs, but the perturbation is gauge dependent. As we saw above, the perturbations in different gauges are defined by:

$$\begin{aligned} \hat{\delta s}(x^\alpha) &\equiv s(\hat{P}) - \bar{s}(\bar{P}) \\ \tilde{\delta s}(x^\alpha) &\equiv s(\tilde{P}) - \bar{s}(\bar{P}) \end{aligned} \tag{7.11}$$

Then, the perturbation δs is just a subtraction between two spacetimes, and for this reason it lives on the background spacetime. We can now relate $\hat{\delta s}$ to $\tilde{\delta s}$:

$$s(\tilde{P}) = s(\hat{P}) + \frac{\partial s}{\partial \hat{x}^\alpha}(\hat{P})(\hat{x}^\alpha(\tilde{P}) - \hat{x}^\alpha(\hat{P})) = s(\hat{P}) - \frac{\partial s}{\partial \hat{x}^\alpha}(\hat{P})\xi^\alpha = s(\hat{P}) - \frac{\partial \bar{s}}{\partial x^\alpha}(\bar{P})\xi^\alpha \tag{7.12}$$

where we approximated $\frac{\partial s}{\partial \hat{x}^\alpha}(\hat{P}) \approx \frac{\partial \bar{s}}{\partial x^\alpha}(\bar{P})$. However, since we are interested in scenarios where the background is homogeneous and time dependent, we have

$\bar{s} = \bar{s}(\eta, x^i) = \bar{s}(\eta)$ only, thus:

$$\frac{\partial \bar{s}}{\partial x^\alpha}(\bar{P})\xi^\alpha = \frac{\partial \bar{s}}{\partial \eta}(\bar{P})\bar{s}'\xi^0 \tag{7.13}$$

so that we get in the end:

$$s(\tilde{P}) = s(\hat{P}) - \bar{s}'\xi^0 \tag{7.14}$$

and the final gauge transformation for scalar perturbations is then:

$$\tilde{\delta s}(x^\alpha) = s(\hat{P}) - \bar{s}'\xi^0 - \bar{s}(\bar{P}) = \hat{\delta s}(x^\alpha) - \bar{s}'\xi^0 \tag{7.15}$$

The other cases, respectively for vectors and tensors in (7.10), are:

$$\begin{aligned} \tilde{\delta w}^\alpha &= \hat{\delta w}^\alpha + \xi^\alpha_{,\beta}\bar{w}^\beta - \bar{w}^\alpha_{,\beta}\xi^\beta \\ \tilde{\delta B}_{\mu\nu} &= \hat{\delta B}_{\mu\nu} - \xi^\rho_{,\mu}\bar{B}_{\rho\nu} - \xi^\sigma_{,\nu}\bar{B}_{\mu\sigma} - \bar{B}_{\mu\nu,\alpha}\xi^\alpha \end{aligned} \tag{7.16}$$

In Cosmological Perturbation Theory the background Universe is described by the flat FRW metric:

$$ds^2 = -dt^2 + a^2(t)d\vec{x}^2 = a^2(\eta)(-d\eta^2 + d\vec{x}^2) \tag{7.17}$$

Therefore, the perturbed spacetime can be written generally as:

$$ds^2 = a^2(\eta)[-(1+2A)d\eta^2 + 2B_i dx^i d\eta + (\delta_{ij} + 2E_{ij})dx^i dx^j] \tag{7.18}$$

where all the new parameters A, B_i and E_{ij} are functions of the spacetime coordinates. It will be extremely useful to perform a scalar-vector-tensor decomposition (SVT) of the perturbations. This just means that any three vector is split into the gradient of a scalar and a divergenceless vector:

$$B_i = \partial_i B + \hat{B}_i \tag{7.19}$$

with $\partial^i \hat{B}_i = 0$. Importantly, in Fourier space the decomposition becomes $B_i = i\hat{k}_i B + \hat{B}_i$, where $\hat{\vec{k}} \equiv \vec{k}/|\vec{k}|$, so that the vector has been split into a piece that is along the same direction of the mode $\vec{k}$ and a piece that is orthogonal to it. In the same manner the following split for tensor perturbations holds:

$$E_{ij} = C\delta_{ij} + \partial_{\langle i}\partial_{j\rangle}E + \partial_{(i}\hat{E}_{j)} + \hat{E}_{ij} \tag{7.20}$$

where

$$\begin{aligned} \partial_{\langle i}\partial_{j\rangle}E &\equiv \left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right)E \\ \partial_{(i}\hat{E}_{j)} &\equiv \frac{1}{2}(\partial_i\hat{E}_j + \partial_j\hat{E}_i) \end{aligned} \tag{7.21}$$

The first term in (7.20) contains the trace of the spatial perturbations $E^i_i = 3C$, while the other ones are traceless. As before, the hatted quantities have vanishing divergence $\partial^i\hat{E}_i = \partial^i\hat{E}_{ij} = 0$, and therefore are the transverse vector and tensor perturbations. In Fourier space, we write the first of (7.21) as $-(\hat{k}_i\hat{k}_j - 1/3\delta_{ij})E \equiv -\hat{k}_{\langle i}\hat{k}_{j\rangle}E$.

What makes the SVT decomposition very powerful and handy is the fact that, at linear order, the Einstein equations for scalars, vectors and tensors don't mix and can therefore be treated completely separately. For the most part we will be interested in scalar perturbation, and at linear order it is consistent to then set vector and tensor perturbations to zero anyway.
Given the transformation of the metric under a gauge transformation $\tilde{x}^\alpha = x^\alpha + \xi^\alpha$ that we found above:

$$\tilde{\delta g}_{\mu\nu} = \delta g_{\mu\nu} - \xi^\rho_{,\mu}\bar{g}_{\rho\nu} - \xi^\sigma_{,\nu}\bar{g}_{\mu\sigma} - \bar{g}_{\mu\nu,0}\xi^0 \tag{7.22}$$

and the form of the perturbations:

$$\delta g_{\mu\nu} = a^2 \begin{pmatrix} 2A & B_i \\ B_i & 2E_{ij} \end{pmatrix} \tag{7.23}$$

if we decompose the gauge transformation as $\xi^0 \equiv T$ and $\xi^i \equiv L^i = \partial^i L + \hat{L}^i$ we get the transformation rules for the perturbations. One can show that:

$$\begin{aligned}
\tilde{A} &= A - T' - \mathcal{H}T \\
\tilde{B} &= B + T - L' \\
\tilde{C} &= C - \mathcal{H}T - \frac{1}{3}\nabla^2 L \\
\tilde{E} &= E - L \\
\tilde{\hat{B}}_i &= \hat{B}_i - \hat{L}'_i \\
\tilde{\hat{E}}_i &= \hat{E}_i - \hat{L}_i \\
\tilde{\hat{E}}_{ij} &= \hat{E}_{ij}
\end{aligned} \tag{7.24}$$

where the conformal Hubble parameter $\mathcal{H} = a'/a$ was introduced. Before carrying on with the calculations, we would like to comment on the whole philosophy at play here. The perturbations are defined as differences between the perturbed spacetime and the background spacetime, so in which spacetime do they live? What we are doing in perturbation theory is building a description of the perturbed spacetime as a set of fields (the perturbations), that live in the background. Thus the background spacetime is the playing field on which we are creating the perturbed spacetime. As we have said, for the same perturbed spacetime, there exist many equivalent descriptions of this kind and they are all related by gauge transformations. This formalism of gauge theory in the background spacetime is similar to gauge theories used in particle physics.

As we briefly mentioned before, the scalar perturbations are the most important ones, since they couple to pressure and density perturbations and follow the gravitational instability, that is overdense regions become more overdense over time, and viceversa for underdense regions. On the other hand, the vector perturbations couple to rotational velocity perturbations in the cosmic fluid, and therefore are expected to decay in an expanding Universe; in this sense we can neglect them. Finally, it is found that tensor perturbations are gauge invariant. These are gravitational waves, automatically in the traceless and transverse gauge, and are very important

since they could have observable effects on CMB anisotropies if strong enough, as mentioned above.

We stressed above that the perturbations are not uniquely defined as a consequence of the gauge invariance of GR. The intrinsic problem stemming from this is the possible appearance of fictitious gauge modes, sham perturbations that can arise from an inconvenient choice of coordinates. As an example, we can consider the transformation of the homogeneous density of the Universe under a local change of coordinates in the time slicing $\eta \to \tilde{\eta} = \eta + \xi^0(\eta, \vec{x})$, given in (7.15):

$$\rho(\eta) = \rho(\tilde{\eta} - \xi^0(\tilde{\eta}, \vec{x})) = \bar{\rho}(\tilde{\eta}) - \bar{\rho}'\xi^0 \tag{7.25}$$

We see that even in a perfectly homogeneous Universe, density perturbations may arise fictitiously if a certain coordinate chart is used rather than another one. Conversely, taking a perturbed Universe, we can remove the perturbations by declaring that we are using coordinates such that $\delta\rho = 0$ (choosing the hypersurface of constant time to coincide with the hypersurface of constant energy density); naturally, in this last case, the perturbations will reappear in some form in the metric.
In general there are two ways one can deal with this problem, which involves a compromise between mathematical simplicity and physical "realness", in the following sense.
We can define the following variables, the "Bardeen variables", which are clever combinations of the metric perturbations that are gauge invariant:

$$\begin{aligned} \Phi_A &\equiv A + \mathcal{H}(B - E') + (B - E)' \\ \Phi_H &\equiv -C + \frac{1}{3}\nabla^2 E - \mathcal{H}(B - E') \end{aligned} \tag{7.26}$$

These, being gauge invariant, would be the "real" spacetime perturbations. An alternative strategy is to fix the gauge and keep track of all the perturbations, both of those in the metric and those in the matter. In this way the math is considerably simpler, and a lot of different gauges are used for different purposes. One that is very important is the Newtonian gauge:

$$B = E = 0 \tag{7.27}$$

The metric becomes then very simple:

$$ds^2 = a^2(\eta)[-(1 + 2\Psi)d\eta^2 + (1 + 2\Phi)\delta_{ij}dx^i dx^j] \tag{7.28}$$

which is the same metric introduced above (6.3), given the identification $\Phi_A = \Psi$ and $\Phi_H = -\Phi$. It is evident that this metric is very useful. First, it is diagonal, so the hypersurfaces of constant time are orthogonal to the wordlines of the observers at rest, but also the induced spatial geometry is isotropic. Finally, this metric resembles the one obtained in the weak-field limit of GR, with Ψ playing the role of gravitational potential. Other useful gauges are the spatially flat gauge ($C = E = 0$) and the synchronous gauge ($A = B = 0$).

7.2 Curvature

The path to the perturbed Einstein equations is now straightforward. In this section we focus on the left hand side, namely the calculation of curvature tensors. The Christoffel's for the metric (7.28) were already calculated in (6.4) and (6.5). The non-zero components of the Ricci tensor instead give, expanded to linear order:

$$\begin{aligned}\delta R_{00} =& -\frac{k^2}{a^2}\Psi - 3\ddot{\Phi} + 3H(\dot{\Psi} - 2\dot{\Phi}) \\ \delta R_{0i} =& 2ik_i(a\dot{\Psi} - \frac{\dot{a}}{a^2}\Phi) \\ \delta R_{ij} =& \delta_{ij}[(2a^2H^2 + a\ddot{a})(2\Phi - 2\Psi) \\ &+ a^2H(6\dot{\Phi} - \dot{\Psi}) + a^2\ddot{\Phi} + k^2\Phi] + k_ik_j(\Phi + \Psi)\end{aligned} \tag{7.29}$$

Whereas the zero-th order components were calculated in (3.5) (with $k = 0$):

$$\begin{aligned}R_{00} &= -3\frac{\ddot{a}}{a} \\ R_{ij} &= \left(\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2\right)\end{aligned} \tag{7.30}$$

These components of the Ricci tensor serve to calculate the perturbation of the Ricci scalar:

$$\delta R = \delta g^{\mu\alpha}R_{\alpha\mu} + g^{\mu\alpha}\delta R_{\alpha\mu} \tag{7.31}$$

Using (7.30) and (7.29) in the above equation yields:

$$\delta R = -12\Psi\left(H^2 + \frac{\ddot{a}}{a}\right) + \frac{2k^2}{a^2}\Psi + 6\ddot{\Phi} - 6H(\dot{\Psi} - 4\dot{\Phi}) + 4\frac{k^2\Phi}{a^2} \tag{7.32}$$

Thus, the curvature, or left hand side, part of the he Einstein equations can be easily calculated. The Einstein equations to first order are, as written in (7.4):

$$\delta G^\mu_\nu = 8\pi G\delta T^\mu_\nu \tag{7.33}$$

where, working with mixed indices (up and down) turns out to simplify the expression for the energy-momentum tensor later. In practice, these are 10 equations, but in linear perturbation theory for scalars, 8 of them are either zero or redundant, decreasing the number of equations to two. The first equation we will use is the time-time component, that is we will need δG^0_0:

$$\delta G^0_0 = (-1 + 2\Psi)\delta R_{00} - \frac{\delta R}{2} \tag{7.34}$$

Using the expressions for the Ricci scalar and tensor, (7.29) and (7.32), one finds:

$$\delta G^0_0 = -6H\dot{\Phi} + 6\Psi H^2 - 2\frac{k^2\Phi}{a^2} \tag{7.35}$$

A second evolution equation for Φ and Ψ comes from the spatial part of the Einstein tensor:

$$\delta G^i_j = \frac{\delta^{ik}(1-2\Phi)}{a^2}\delta R_{kj} - \frac{\delta^i_j}{2}\delta R \tag{7.36}$$

Plugging the perturbed Ricci tensor and scalar into (7.36) we find the following neat expression:

$$\delta G^i_j = F[\Phi,\Psi]\delta^i_j + \frac{k^i k_j(\Phi+\Psi)}{a^2} \tag{7.37}$$

where F is some functional of the potentials, and it contributes to the trace of the spatial Einstein tensor. To avoid dealing with all these terms proportional to δ_{ij}, one in general extracts the longitudinal and traceless part of G^i_j, multiplying it by $\hat{k}_i\hat{k}^j - (1/3)\delta^i_j$. This procedure readily yields:

$$\left(\hat{k}_i\hat{k}^j - \frac{1}{3}\delta^i_j\right)\delta G^i_j = \frac{2}{3a^2}k^2(\Phi+\Psi) \tag{7.38}$$

7.3 Matter and energy

The right hand side of the Einstein equations (7.33) can be calculated using the general expression for the energy momentum tensor given in (6.2). In particular the time-time component is:

$$T^0_0(\mathbf{x},t) = -\sum_s g_s \int \frac{d^3p}{(2\pi)^3}\, E_s(p) f_s(\mathbf{p},\mathbf{x},t) \tag{7.39}$$

where the sum runs over the species with spin s, degeneracy factor g_s, energy $E_s(p) = \sqrt{p^2+m_s^2}$ and distribution function f_s. The first order portion of the energy-momentum tensor is derived using the first order pieces of the distribution functions, which contain all the matter perturbations introduced in the last chapter, such as bulk velocities, density contrasts, temperature anisotropies etc... . With regards to dark matter and baryons, we can just use $E_s(p) \simeq m_s$, so that their contribution to T^0_0 amounts to adding a factor $-mn(t,\mathbf{x})$, with n being the number density. Thus, we have:

$$T^0_0|_{s=b,c} = -\rho_s(1+\delta_s) \tag{7.40}$$

For photons, we can use the expansion found in (6.30) and write:

$$T^0_0|_{s=\gamma} = -2\int \frac{d^3p}{(2\pi)^3}\left(f^{(0)} - p\frac{\partial f^{(0)}}{\partial p}\Theta\right) \tag{7.41}$$

The first factor gives the radiation energy density ρ_γ. The second factor can be handled by a first integration over the angle, which picks out the monopole Θ_0, and then by doing an integration by parts over p. The result is the following:

$$T^0_0|_{s=\gamma} = -\rho_\gamma[1+4\Theta_0] \tag{7.42}$$

The factor of 4 is a result of integration by parts, but should be clear from a physical standpoint. The temperature perturbation is in fact $\Theta = \delta T/T$, but since for radiation we have $\rho_\gamma \propto T^4$, the fractional density perturbation is only $\delta\rho_\gamma/\rho_\gamma = 4\delta T/T$.

Finally, the calculation for the neutrinos follows exactly the same steps as that for photons (if we assume they have no mass), so:

$$T^0_0|_{s=\nu} = -\rho_\nu[1 + 4\mathcal{N}_0] \tag{7.43}$$

If instead one assumes $m_\nu \neq 0$, the integral over the momentum cannot be solved in closed form. In the next chapter, we will address the evolution of perturbations under the assumption of zero neutrino masses. Additionally, we will discuss the impact of non-zero neutrino masses based on numerical solutions.
The time-time component of the Einstein equations then reads:

$$-3H\dot{\Phi} + 3\Psi H^2 - \frac{k^2}{a^2}\Phi = -4\pi G[\rho_c\delta_c + \rho_b\delta_b + 4\rho_\gamma\Theta_0 + 4\rho_\nu\mathcal{N}_0] \tag{7.44}$$

It is also useful to write the equation in conformal time:

$$k^2\Phi + 3\mathcal{H}\left(\Phi' - \Psi\mathcal{H}\right) = 4\pi G a^2[\rho_c\delta_c + \rho_b\delta_b + 4\rho_\gamma\Theta_0 + 4\rho_\nu\mathcal{N}_0] \tag{7.45}$$

For the second Einstein equation, we need to equate (7.38) to the traceless and longitudinal part of the energy-momentum tensor. From (6.2) again we have:

$$T^i_j(\mathbf{x}, t) = \sum_s g_s \int \frac{d^3p}{(2\pi)^3}\,\frac{p^i p_j}{E_s(p)} f_s(\mathbf{x}, \mathbf{p}, t) \tag{7.46}$$

The projection operator brings this expression to:

$$\left(\hat{k}_i\hat{k}^j - \frac{1}{3}\delta^i_j\right) T^i_j = \sum_s g_s \int \frac{d^3p}{(2\pi)^3}\,\frac{p^2\mu^2 - (1/3)p^2}{E_s(p)} f_s(\mathbf{p}) \tag{7.47}$$

where we have used the definition of μ via $\hat{\mathbf{k}}\cdot\mathbf{p} = kp$. Immediately, one can spot the combination $\mu^2 - 1/3$ as proportional to the second Legendre polynomial, more precisely equal to $(2/3)\mathcal{P}_2(\mu)$. Therefore, the integral above picks out the quadrupole of the distribution function. However, since the zero-th order distribution $f^{(0)}$ has no quadrupole, this source term is only non-zero for photons and neutrinos. For example, for photons one has:

$$\begin{aligned} -2\int \frac{dpp^2}{2\pi^2}\, p^2\frac{\partial f^{(0)}}{\partial p} \int_{-1}^{+1}\frac{d\mu}{2}\,\frac{2\mathcal{P}_2(\mu)}{3}\Theta(\mu) &= 2\frac{2\Theta_2}{3}\int \frac{dpp^2}{2\pi^2}\, p^2\frac{\partial f^{(0)}}{\partial p} \\ &= -\frac{8}{3}\rho_\gamma\Theta_2 \end{aligned} \tag{7.48}$$

This component of the energy-momentum tensor is called the anisotropic stress. Non-relativistic particles do not contribute to the anisotropic stress, because of the suppression acted the factor $p/E_s(p)$. Thus, we can write down the second Einstein equation:

$$k^2(\Phi + \Psi) = -32\pi G a^2[\rho_\gamma\Theta_2 + \rho_\nu\mathcal{N}_2] \tag{7.49}$$

This equation immediately tells us that the two potentials are in fact the same (except for a sign difference), if the radiation content of the Universe does not carry a considerable quadrupole moment. In practice, the term Θ_2 contributes negligibly

during the time of tight-coupling, as was laid out in the last chapter. Only neutrinos have an appreciable quadrupole moment before recombination, because of their early decoupling. Equation (7.49) is technically a constraint equation, as it does not contain any time derivatives of the perturbations, that is Ψ and Φ do not represent propagating degrees of freedom, which in GR only are the tensor modes. Equations (7.45) and (7.49) are the sought for Einstein equations for the metric perturbations Ψ and Φ. The next chapter is finally concerned with solving the whole system of coupled differential equations in the perturbation variables for the metric and the energy-matter.

Chapter 8

The Linear Evolution of Matter Perturbations

Once the equations for the perturbation variables are set, one can in principle solve them to find their evolution in a given background expanding spacetime. In this chapter, particular attention is given to the evolution of dark matter perturbations, that is the density contrast δ_c and the bulk velocity $\mathbf{u}_c$, which are only coupled to other perturbations via the metric. For this reason, an in depth analysis of the photon perturbations is not necessary and will not be carried out in these lectures. For the purposes of this chapter, all one needs to know with regards to the photon perturbations is that they are not coupled to the potentials Φ and Ψ after recombination, so that dark matter is not coupled to them in this era, and that before recombination they are completely determined by their monopole and dipole.
The goal of this chapter is to give a prediction for an observable called the "linear matter power spectrum" of density perturbations. This, as we shall see, is the most simple of quantities that can be predicted theoretically and tested against observations. The power spectrum is defined as:

$$\langle \delta_m(\mathbf{k},\eta)\delta_m(\mathbf{k}',\eta)\rangle = (2\pi)^3\delta_D^{(3)}(\mathbf{k}-\mathbf{k}')P_m(k,\eta) \tag{8.1}$$

where the average symbols represent an ensemble average. While the physical significance of the power spectrum will be discussed in more detail later, it is immediately clear that it represents the variance of the density field. Matter, on small enough scales, becomes non-linear and the result for the linear matter power spectrum is not valid anymore. The regime where δ_c cannot be considered a small variable will be treated in the next chapter.

8.1 Overview

As we laid out in the last few chapters, one needs to solve the Einstein and Boltzmann equations, which form a system of coupled partial differential equations. For organization purposes, in the following we will lay them out (all in Fourier space and in conformal time). Starting with the Einstein equations, these can be written as:

$$\begin{aligned} k^2\Phi + 3\mathcal{H}\left(\Phi' - \Psi\mathcal{H}\right) &= 4\pi G a^2[\rho_m\delta_m + 4\rho_r\Theta_{r,0}] \\ k^2(\Phi+\Psi) &= -32\pi G a^2\rho_r\Theta_{r,2} \end{aligned} \tag{8.2}$$

The subscript m indicates all matter, such as baryons and dark matter, and the subscript r all radiation, that is neutrinos and photons. More precisely,

$$\begin{aligned} \rho_m \delta_m &= \rho_c \delta_c + \rho_b \delta_b \quad \rho_r \Theta_{r,0} = \rho_\gamma \Theta_0 + \rho_\nu \mathcal{N}_0 \\ \rho_m u_m &= \rho_c u_c + \rho_b u_b \quad \rho_r \Theta_{r,1} = \rho_\gamma \Theta_1 + \rho_\nu \mathcal{N}_1 \end{aligned} \tag{8.3}$$

The other Einstein equations are either zero or redundant, that is they offer no new information on the potentials. Next, we collect the equations we have derived for photons, dark matter, baryons and neutrinos. Starting with photons, they obey:

$$\Theta' + ik\mu\Theta = -\Phi' - ik\mu\Psi - \tau' \left[\Theta_0 - \Theta + \mu u_b - \frac{1}{2}\mathcal{P}_2(\mu)\Pi \right] \tag{8.4}$$

where:

$$\Pi = \Theta_2 + \Theta_{P,2} + \Theta_{P,0} \tag{8.5}$$

The last term added to the equation, which was not there in our derivation, needs some kind of explanation. First, it is proportional to the second Legendre polynomial, and the term $\mathcal{P}_2(\mu)\Theta_2/2$ has to do with the angular dependence of Compton scattering, which we neglected. The other parts of Π represent the fact that the temperature field is coupled to the polarization field Θ_P, as was hinted at before; we will not give a derivation of the polarization terms in these notes. Next, the dark matter perturbations obey the following equations:

$$\begin{aligned} \delta_c' + iku_c &= -3\Phi' \\ u_c' + \mathcal{H}u_c &= -ik\Psi \end{aligned} \tag{8.6}$$

The baryon equations are similar, but with an interaction term added:

$$\begin{aligned} \delta_b' + iku_b &= -3\Phi' \\ u_b' + \mathcal{H}u_b &= -ik\Psi + \frac{\tau'}{R}[u_b + 3i\Theta_1] \end{aligned} \tag{8.7}$$

l

where the ratio of photons to baryon density was replaced by R, defined as:

$$\frac{1}{R(\eta)} = \frac{4\rho_\gamma(\eta)}{3\rho_b(\eta)} \tag{8.8}$$

Finally, neutrinos obey an equation similar to that of photons:

$$\mathcal{N}' + ik\mu\frac{p}{E_\nu(p)}\mathcal{N} - \frac{\mathcal{H}}{ap}\frac{\partial}{\partial p}\mathcal{N} = -\Phi' - ik\mu\frac{E_\nu(p)}{p}\Psi \tag{8.9}$$

These equations precisely map out the evolution of matter perturbations and their interactions. Gravitational instability is the culprit of structure formation, as matter tends to flow towards initially slightly overdense regions, and as a direct consequence flow away from underdense regions. Initial overdensities were very small (of the order of 10^{-4}), but grew to be the structures we see today in the Universe.
As far as dark matter is concerned, there are only two forces at play: the attractive force of gravity and the expansion of the Universe, which have opposing effects. Gravity pulls matter in at a rate that can only be consistent with the expansion of

spacetime, which is the reason why, as we will see below, dark matter perturbations can only grow significantly in the matter era as opposed to the radiation era, when the expansion is so fast that it overpowers the pull of gravity. Once dark matter is free to cluster, there is no stopping the process, as it is a non-interacting species. The final result is the formation of dark matter virialized objects, among which dark matter halos, into which the baryons fall and start the process of galaxy formation. Bigger and bigger objects would be created as the Universe evolves were it not for dark energy, which speeds up the Universe to such an extent that it completely halts structure formation.
When it comes to baryons, an additional force due to photon pressure is at play, given the tight coupling between those two components. A consequence of this phenomenon is that, before recombination, baryon perturbation oscillate in the form of sound waves. Gravity tends to cluster baryons but, as opposed to dark matter, the free fall is halted at some stage because of photon pressure, which re-expands the overdense region. After recombination, the baryons are free to cluster under the help of dark matter, which in the mean time has already formed potential wells, in the form of dark matter halos. In other words, the existence of dark matter is a prerequisite for the existence of galaxies.
In this chapter, both the super-(particle)horizon ($k\eta \ll 1$) and the sub-horizon ($k\eta \gg 1$) versions of gravitational evolution will be treated.

8.2 Inflation and Initial Conditions

As for any system of coupled differential equations, unique solutions can be found only by specifying initial conditions. In principle, for scalar perturbations this means defining initial density and velocity perturbations for each species.
The formalism of inflation, as was introduced above, besides solving the flatness and horizon problems, offers a framework with which to set initial conditions. In fact, cosmological perturbation theory starts at some time after inflation when all scales are outside the horizon, where they have been pushed by a period of accelerated expansion due to a scalar field ϕ, the inflaton. This field naturally predicts a situation where the Universe is homogeneous with some amounts of perturbations. This, when the field ϕ is quantized, is achieved through quantum fluctuations. In other words, the inflaton can be decomposed into a zero-th order homogeneous part and its perturbations:

$$\phi(\mathbf{x}, t) = \bar{\phi}(t) + \delta\phi(\mathbf{x}, t) \tag{8.10}$$

as was done in (5.75), but now we use the notation $\phi_0 \equiv \bar{\phi}$. At the end of inflation, the Universe is bound to be homogeneous: since the energy density in $\bar{\phi}$ is the same everywhere in the Universe, the resulting temperature is also uniform. The introduction of perturbations through $\delta\phi$ instead implies that inflation ends at slightly different times in different locations in space, leading to density perturbations when reheating takes place. Therefore, all perturbations are "the same", in the sense that they are all related to quantum fluctuations of the same object, and initial conditions will be given such that $\delta_{c,b} \sim \delta\phi$ and $u_{c,b} \sim \delta\phi$. Indeed, in cosmological nomenclature, inflation is said to generate "adiabatic perturbations", that is different patches of the Universe have different overdensities, but the fractional density perturbations

are the same for all species:

$$\frac{\delta\rho_s}{\rho_s} = \frac{\delta\rho}{\rho} \tag{8.11}$$

The goal, now, is to derive how the fluctuations $\delta\phi$ are related to the value of the density contrasts at early times, that is outside the horizon. Because of the Einstein equations, instead of δ, we can specify initial conditions in terms of Ψ (or Φ), since outside the horizon they are related (see below). Since the Universe is a stochastic process (see below, the discussion around the power spectrum), it then all comes down to determining how the power spectrum of the inflaton perturbation $P_{\delta\phi}$ is converted into a power spectrum for the potential P_Ψ.
The way it is usually done in the literature, is that one considers the evolution of $\delta\phi$ in an expanding homogeneous Universe, that is neglecting its coupling with the metric perturbations. This will be justified below by noticing that, until a mode moves outside the horizon, Ψ is negligibly small. Once it is far outside the horizon, this no longer holds, but a linear combination of Ψ and $\delta\phi$ is actually conserved through horizon crossing. This will allow us to convert the initial power spectrum for $\delta\phi$ into a final spectrum for Ψ.
Proceeding in the said manner, consider the evolution of $\delta\phi$ through the $\nu = 0$ energy-momentum conservation equation $\nabla_\mu T^{\mu\nu} = 0$. The metric is taken to be the flat FRW metric, whereas the energy momentum tensor is (5.42), but expanded to first order. The $\nu = 0$ equation then reads:

$$0 = \frac{\partial}{\partial t}\delta T^0_0 + ik_i\delta T^i_0 + 3H\delta T^0_0 - H\delta T^i_i \tag{8.12}$$

The perturbations in the energy-momentum tensor can be easily calculated, and read:

$$\begin{aligned}
\delta T^i_0 &= \frac{ik_i}{a^3}\bar{\phi}'\delta\phi \\
\delta T^0_0 &= -\frac{\bar{\phi}'\delta\phi'}{a^2} - \frac{\partial V}{\partial\phi}\delta\phi \\
\delta T^i_j &= \delta_{ij}\left[\frac{\bar{\phi}'\delta\phi'}{a^2} - \frac{\partial V}{\partial\phi}\delta\phi\right]
\end{aligned} \tag{8.13}$$

Therefore, the conservation equation (8.12) becomes:

$$\left(\frac{1}{a}\frac{\partial}{\partial\eta} + 3H\right)\left(-\frac{\bar{\phi}'\delta\phi'}{a^2} - \frac{\partial V}{\partial\phi}\delta\phi\right) - \frac{k^2}{a^3}\bar{\phi}'\delta\phi - 3H\left(\frac{\bar{\phi}'\delta\phi'}{a^2} - \frac{\partial V}{\partial\phi}\delta\phi\right) = 0 \tag{8.14}$$

Carrying out the derivatives and multiplying everything by a^3 gives:

$$-\bar{\phi}'\delta\phi'' + \delta\phi'(-\bar{\phi}'' - 4aH\bar{\phi}' - a^2V_{,\phi}) + \delta\phi(-a^2V_{,\phi\phi}\bar{\phi}' - k^2\bar{\phi}') = 0 \tag{8.15}$$

The term $V_{,\phi\phi}$ is typically small, as it is proportional to the slow roll variables (see for instance (5.64)), so it can be neglected. In addition, the second equation in (5.52) (written in conformal time), implies that the factor in parenthesis multiplying $\delta\phi'$ is equal to $-2aH\bar{\phi}'$. After dividing by $-\bar{\phi}'$, we are then left with:

$$\delta\phi'' + 2aH\delta\phi' + k^2\delta\phi = 0 \tag{8.16}$$

Equation (8.16) can be shown to be equivalent to a harmonic oscillator-type equation, so that after a suitable quantization of the inflaton, its vacuum expectation value can be calculated using standard quantum mechanics techniques. Defining the power spectrum as in (8.1):

$$\langle \delta\phi(\mathbf{k},\eta)\delta\phi(\mathbf{k}',\eta)\rangle = (2\pi)^3\delta_D^{(3)}(\mathbf{k}-\mathbf{k}')P_{\delta\phi}(k,\eta) \tag{8.17}$$

one easily finds:

$$P_{\delta\phi} = \frac{H^2}{2k^3} \tag{8.18}$$

This is the primordial power spectrum, and it is the starting point for the calculation of the late-time matter power spectrum, as we will see below.
The next task is to relate the two variables $\delta\phi$ and ψ. As we said above, since Ψ cannot be neglected once the given mode moves outside the horizon, the trick is to find a linear combination of the two perturbations that is conserved through horizon crossing. This combination is then evaluated at horizon crossing, and is then matched to its value after inflation ends, such that the resulting equation will be of the form $\Psi \propto \delta\phi$, which is the sought for relation. Let us first rewrite the $\nu = 0$ conservation equation for the energy-momentum tensor, but this time with the metric perturbations. It is easy to find:

$$0 = \frac{\partial}{\partial t}\delta T_0^0 + ik_i\delta T_0^i + 3H\delta T_0^0 - H\delta T_i^i + 3(\rho+P)\dot{\Psi} = 0 \tag{8.19}$$

where ρ and P are the zero-th order energy density and pressure, and $\dot{\Phi}$ was replaced by $-\dot{\Psi}$. Therefore, as long as the last term is significantly smaller than the others, we were justified in neglecting it above. This is in fact the case, since the Einstein equations imply $\Psi \sim \delta T_0^0/\rho$ (see later), such that all terms except the last one in (8.19) are of order $\rho\Psi$. On the other hand, slow-roll inflation implies that $|\rho + P| \ll \rho$, thus suppressing the last term.
Next, we define the "curvature perturbation" $\mathcal{R}$ in the following way:

$$\mathcal{R}(\mathbf{k},\eta) = \frac{ik_i\delta T_0^i(\mathbf{k},\eta)a^2H(\eta)}{k^2(\rho+P)(\eta)} - \Psi(\mathbf{k},\eta) \tag{8.20}$$

The name of the variable comes from the fact that, in the comoving and uniform density gauges, they reduce to the intrinsic curvature of spatial slices. This variable is important because it can be proven to be conserved when the mode moves outside the horizon. During inflation, we then know that Ψ is negligible compared to the first term, and also that $p + \rho = (\bar{\phi}'/a)^2$, which follows from (5.43). Therefore, during inflation, the curvature perturbation takes the value:

$$\mathcal{R} = -\frac{aH}{\bar{\phi}'}\delta\phi \quad \text{During inflation} \tag{8.21}$$

After inflation ends, at the beginning of the era of radiation domination, we have that $ik_i\delta T_0^i = -4\rho_r\Theta_1/a$, following the definition of the energy-momentum tensor in (6.2). Since the pressure equals one third of the energy density, we have that the

post-inflationary curvature perturbation is:

$$\mathcal{R} = -\frac{3aH\Theta_1}{k} - \Psi = -\frac{3}{2}\Psi \quad \text{Post inflation} \tag{8.22}$$

where the second equality will be derived below, see (8.39). Since the variable is conserved, we can relate the value of Ψ right after inflation to the inflaton fluctuations $\delta\phi$ at horizon crossing, that is:

$$\Psi\Big|_{\text{post inflation}} = \frac{2}{3}aH\frac{\delta\phi}{\bar{\phi}'}\Big|_{\text{horizon crossing}} \tag{8.23}$$

Equivalently, using the definition of power spectrum in (8.17), the relation between power spectra read:

$$P_\Psi(k)\Big|_{\text{post inflation}} = \frac{4}{9}\left(\frac{aH}{\bar{\phi}'}\right)^2 P_{\delta\phi}(k)\Big|_{aH=k} = \frac{2}{9k^3}\left(\frac{aH^2}{\bar{\phi}'}\right)^2\Big|_{aH=k} \tag{8.24}$$

where in the second line the inflaton's power spectrum (8.18) was used. Note that, because of the equality $\Psi = -\Phi$, we have that $P_\Phi = P_\Psi$.

Early-times Einstein-Boltzmann equations

At this stage, it only remains to connect the initial (after inflation) metric perturbation Ψ to the other matter-energy perturbation variables. Since inflation sources adiabatic perturbations, the initial conditions will be relatively simple, as one expects all density contrasts and velocities to be determined by the same perturbation Ψ, which is in turn related to $\mathcal{R}$, and therefore $\delta\phi$ through (8.21). It is in this sense that all structures in the Universe, given the theory of single field inflation, are the result of quantum fluctuations created during the early Universe.
First, consider the Boltzmann equations (8.4) through (8.9) at very early times, just after inflation. All the modes are outside the particle horizon (at least those stretching out to our observable Universe today) and therefore the limit of interest for every k is $k\eta \ll 1$, or equivalently $k/aH \ll 1$. Consider first the equation governing photons:

$$\Theta' + ik\mu\Theta = -\Phi' - ik\mu\Psi - \tau'\left[\Theta_0 - \Theta + \mu u_b - \frac{1}{2}\mathcal{P}_2(\mu)\Pi\right] \tag{8.25}$$

Clearly, the first term, of order Θ/η, overpowers the terms proportional to k, that is $k\Psi$ and $k\Theta$. In addition, the terms u_b, Π can be neglected, since all perturbations of interest have wavelengths k^{-1} much larger than the distance over which causality operates. An observer, who sees only photons coming from within his horizon, will necessarily see a uniform sky at the end of inflation. In the same manner, higher multipoles of Θ can be neglected, such that $\Theta \sim \Theta_0$. The neutrino and photon equation reduce, at early times, to:

$$\begin{aligned} \Theta_0' + \Phi' &= 0 \\ \mathcal{N}_0' + \Phi' &= 0 \end{aligned} \tag{8.26}$$

The same reasonings can be applied to the Boltzmann equations for cold dark matter (8.6) and baryons (8.7):

$$\begin{aligned} \delta_c' &= -3\Phi' \\ \delta_b' &= -3\Phi' \end{aligned} \tag{8.27}$$

The equations for CDM and baryons are the same at early times because the modes are outside the horizon. Since causal physics cannot operate, the two species cannot communicate other than through gravity, which is the only relevant interaction at such length scales.
As far as the Einstein equations are concerned, consider the first of (8.2). The term proportional to k^2 can be neglected and, because of radiation domination, the two matter terms on the right-hand side are sub-leading compared to the radiation terms. Therefore, we have:

$$3\mathcal{H}(\Phi' - \mathcal{H}\Psi) = 16\pi G a^2 \rho_r \Theta_{r,0} \tag{8.28}$$

In the radiation domination era, $a \propto \eta$, so $\mathcal{H} = 1/\eta$, thus:

$$\frac{\Phi'}{\eta} - \frac{\Psi}{\eta^2} = \frac{2}{\eta^2}\Theta_{r,0} \tag{8.29}$$

where the Friedmann equation was used. Multiplying by η^2 yields:

$$\Phi'\eta - \Psi = 2\Theta_{r,0} \tag{8.30}$$

Differentiating with respect to η and using the relations (8.26) to eliminate the monopoles, we have:

$$\Phi''\eta + \Phi' - \Psi' = -2\Phi' \tag{8.31}$$

At this point, the second Einstein equation can be used to find a relation between the two potentials. For the moment, neglecting the higher-order moments of the neutrino and photon distributions (which is a good approximation, as argued above), we have that $\Phi = -\Psi$. Then (8.31) becomes:

$$\Phi''\eta + 4\Phi' = 0 \tag{8.32}$$

The ansatz $\Phi = \eta^p$ easily leads to two solutions: $p = 0$ and $p = -3$. The $p = -3$ mode is decaying, so we can neglect it, as it will quickly die out as the Universe evolves. The $p = 0$ mode, used in the relation (8.30), implies that the gravitational potential and the total radiation overdensities are related at initial times:

$$\Phi = 2\Theta_{r,0} \tag{8.33}$$

For adiabatic perturbations, on the other hand we have:

$$\Theta_0(\mathbf{k}, \eta_i) = \mathcal{N}_0(\mathbf{k}, \eta_i) \tag{8.34}$$

This leads to:

$$\Phi(\mathbf{k}, \eta_i) = 2\Theta_0(\mathbf{k}, \eta_i) \tag{8.35}$$

For the initial conditions regarding cold dark matter we have, combining the first in (8.26) with the first in (8.27) gives:

$$\delta_c(\mathbf{k},\eta) = 3\Theta_0(\mathbf{k},\eta) + \text{const}(\mathbf{k}) \tag{8.36}$$

The baryon overdensity satisfies an identical equation, with the constant being the same due to the perturbations being adiabatic. This constant is in fact zero, since adiabaticity leads to a uniform matter-to-photon ratio, given by:

$$\frac{n_c}{n_\gamma} = \frac{\bar{n}_c}{\bar{n}_\gamma}\left[\frac{1+\delta_c}{1+3\Theta_0}\right] \tag{8.37}$$

where the factor 3 appears because we are considering fluctuations in the number density for photons, not in the energy density. The prefactor is constant in both space and time, therefore, for the left-hand side to be uniform, the term inside the brackets must be independent of space. Since this term is linearized to $1+\delta_c-3\Theta_0$, the perturbations must sum to zero, that is:

$$\delta_c = \delta_b = 3\Theta_0 \tag{8.38}$$

We also need the initial conditions for the velocities and dipole moments of matter and radiation. These can be shown to satisfy:

$$\Theta_1(\mathbf{k},\eta_i) = \eta_1(\mathbf{k},\eta_i) = \frac{iu_b(\mathbf{k},\eta_i)}{3} = \frac{iu_c(\mathbf{k},\eta_i)}{3} = -\frac{k}{6aH}\Phi(\mathbf{k},\eta_i) \tag{8.39}$$

With these relations, the second equality in (8.22) is derived. To summarize the initial conditions for the overdensities and the monopoles, we write them in a similar way to (8.39):

$$\delta_c(\mathbf{k},\eta_i) = \delta_b(\mathbf{k},\eta_i) = 3\Theta_0(\mathbf{k},\eta_i) = 3\eta_0(\mathbf{k},\eta_i) = -\frac{3}{2}\Psi \tag{8.40}$$

The initial conditions (8.40) and (8.39) were the goal of this section. Notice that they are in the form that was suggested at the beginning of the discussion, where we postulated that $\delta \sim \delta\phi$ and $u \sim \delta\phi$ at initial times. These are clearly reflected in the two equations above, since the potentials, related to each other, are proportional to the curvature perturbation $\mathcal{R}$ via (8.22), which is sourced by $\delta\phi$ itself via (8.21).

8.3 Observations and the Matter Power Spectrum

In this section we would like to offer a quantitative analysis of the (late time) matter power spectrum. It should be now clear that, observations of large-scale structures in the Universe, such as galaxy surveys, weak lensing measurements, cluster counts etc., have revealed the picture of a homogeneous and isotropic Universe, pervaded by structures and inhomogeneities at smaller scales. The linear evolution of these inhomogeneities from inflation to the present day is the subject of this chapter. The most basic of observables that quantifies the degree of inhomogeneity in the Universe is the matter power spectrum. In fact, since the contrast fields are defined with respect to the mean itself, we trivially have $\langle\delta_m\rangle = 0$. Then, the next statistically relevant quantity is:

$$\langle\delta_m(\mathbf{x},t)\delta_m(\mathbf{x}',t)\rangle = \xi(t,\mathbf{x},\mathbf{x}') \tag{8.41}$$

where the object appearing on the right is called the correlation function. Because of homogeneity and isotropy, we can rewrite this as:

$$\xi(r) = \langle \delta_m(\mathbf{x}) \delta_m(\mathbf{x'}) \rangle \tag{8.42}$$

where $r = |\mathbf{x} - \mathbf{x'}|$ and the time t is implied. This is fundamental, because it asks the question: if I have a certain over(under)-density here $\delta_m(\mathbf{x})$, is it correlated to an over(under)-density over there $\delta_m(\mathbf{x'})$? This is what we wanted of course, that is $\xi(r)$ measures of the clumpiness of the Universe at a scale r. Working in Fourier space is where the power spectrum comes into play. Let's first transform the δ_m:

$$\delta_m(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \delta_m(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{8.43}$$

This, as is known, is just a decomposition of the field into its waves, thus $|\delta(\mathbf{k})|^2$ is the amplitude of a certain wave-number, i.e. how important that wave-number is in the decomposition. In fact, we may give the following pseudo-definition that, although not rigorous, is the basis of the idea behind the matter power spectrum $P_m(k)$:

$$|\delta(\mathbf{k})|^2 \sim P_m(k) \tag{8.44}$$

In other words, the power spectrum indicates how much "power" resides at a certain scale $\lambda \sim 1/k$. What does this mean? The Fourier transform "divides" 3D space into pieces, which can either be big (small k) or small (large k). Therefore, if, say, $P_m(k)$ has a peak at a certain $k = k_\star$, this means that the perturbations with wavelengths, i.e. scales, $\lambda_\star \sim 1/k_\star$ are very important, which ultimately results in the Universe being highly correlated at $\lambda_\star$. Even though this line of thought gives a rather intuitive explanation, it is only roughly true, leaving behind some details which need some clarification. Clearly, equations (8.42) and (8.44) are very alike, and in fact we can be more precise (although the mathematics will become more involved, the physical intuition given above remains valid). At a glance, we might think of the power spectrum as the Fourier transform of the correlation function. This is exactly what is done traditionally, and may even serve as a definition:

$$\begin{aligned}
\langle \delta_m(\mathbf{k}) \delta_m(\mathbf{k'}) \rangle &= \int d^3x d^3x' e^{-i\mathbf{k}\cdot\mathbf{x}} e^{i\mathbf{k'}\cdot\mathbf{x'}} \langle \delta_m(\mathbf{x}, t) \delta_m(\mathbf{x'}, t) \rangle \\
&= \int d^3r d^3x' e^{-i\mathbf{k}\cdot\mathbf{r}} e^{-i(\mathbf{k}-\mathbf{k'})\cdot\mathbf{x'}} \xi(r) \\
&= (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k'}) \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(r) \\
&\equiv (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k'}) P_m(k)
\end{aligned} \tag{8.45}$$

The last line clearly isn't that much different from (8.44). This equation is to be compared to (8.1), where the subscript m is now used to indentify all matter, instead of only CDM. We have then defined:

$$P_m(k) = \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(r) \tag{8.46}$$

On the other hand then:

$$\begin{aligned}\xi(r) &= \langle \delta_m(\mathbf{x})\delta_m(\mathbf{x}+\mathbf{r})\rangle \\ &= \int \frac{d^3k}{(2\pi)^3}\, P_m(k)e^{i\mathbf{k}\cdot\mathbf{r}} = \frac{1}{2\pi^2}\int_0^\infty dk\, k^2 P_m(k)\frac{\sin(kr)}{kr}\end{aligned} \tag{8.47}$$

Practically, even though it is $\xi(r)$ that measures the clumpiness of the Universe in real space, the matter power spectrum is the quantity that's mathematically easier to handle.
Now, one might ask what these averages $\langle \dots \rangle$ mean, and why we are even taking them. The answer lies in one critical observation, namely that the Universe is an intrinsically stochastic process. Stochasticity is a property of those systems for which it is impossible to exactly predict the evolution, because they contain a fundamental element of chance. This randomness could be attributed to an incomplete knowledge of the initial conditions, or simply to measurement errors. These systems are described through random variables. A random variable is determined by a set of numbers (its values) and a probability measure defined over this set. Then one considers a large (ideally infinite) collection of identical systems (an ensemble) and builds ensemble averages from there, in order to define the statistics of that random variable (becasue of fundamental stochasticity, statistics are all we have).
A typical example is that of a coin toss. Here the randomness is obviously attributed to our ignorance about initial conditions. A random variable that could be defined is Y, a variable that models a 1\$ payoff for a successful bet on heads. Its values are twofold, 1 and 0. The probability measure defined on this set is clearly $1/2$ and $1/2$ respectively. Now, we might ask the following question: what is the probability of having 20\$ after 50 throws? The answer is: toss the coin 50 times a large (again, ideally infinite) amount of times, and average over the results, i.e. take the ensemble average.
Our case is fundamentally the same. Clearly, to know the initial conditions of the Universe, we would need an infinite set of values: $\delta_m(\mathbf{x}, t_i)$ for every $\mathbf{x}$, which is obviously impossible. Now, we are dealing with random fields, like $\delta_m(x)$, instead of random variables, but the former are roughly infinite and continuous collections of random variables (one for each point x). If we now ask the probability of correlations in matter densities as a function of scale, i.e. we ask for $\xi(r)$, the line of thinking is clear. We take multiple realizations of the Universe, and then we average over all of these against the probability distribution function which defines the random field; this sequence defines the "observational" two-point function $\xi_{\rm obs}$:

$$\xi_{\rm obs}(r) = \int \mathcal{D}\delta_m \mathbb{P}[\delta_m]\delta_m(\mathbf{x}, t)\delta_m(\mathbf{x'}, t) \tag{8.48}$$

Again, since the Universe is a fundamentally stochastic process, statistics are all we can hope to obtain. Next, then, one has to specify what the probability measure of the random variables (in total, a field) is. There is good reason to believe that it is a Gaussian:

$$\mathbb{P}[\delta_m]d\delta_m = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left(-\frac{\delta_m^2}{2\sigma^2}\right)d\delta_m \tag{8.49}$$

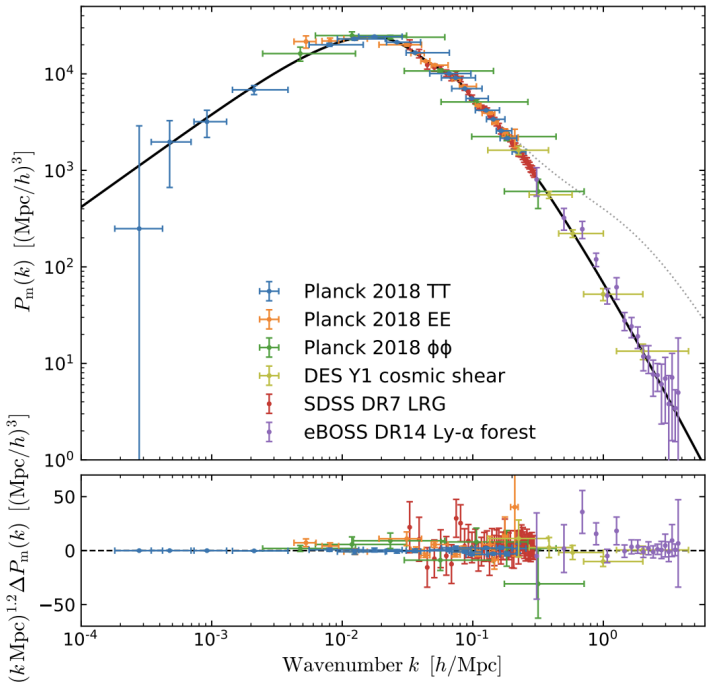

FIGURE 8.1: Matter power spectrum

where $\sigma^2 = \xi(0)$. Gaussian random fields are particularly simple and convenient, since, as long as the evolution of the δ_m is linear, the statistics remain Gaussian. Importantly, this means that even initially they were Gaussian. This is in fact a prediction of single field inflation. In addition, the two point correlation function specifies completely all other N-point functions, as the joint probability reads:

$$\mathbb{P}[\delta_m(\mathbf{x}_1), \delta_m(\mathbf{x}_2), \ldots, \delta_m(\mathbf{x}_N)] \propto \frac{1}{\sqrt{\det(\xi_{ij})}} \exp\left(-\frac{1}{2}\delta_m(\mathbf{x}_i)\xi_{ij}^{-1}\delta_m(\mathbf{x}_j)\right) \tag{8.50}$$

where $\xi_{ij} \equiv |\mathbf{x}_i - \mathbf{x}_j|$. There is a slight problem with this line of reasoning however. It has to do with the obvious fact that we don't possess any number of copies of Universes to average on, let alone an infinite amount. For this reason, in order to test theoretical predictions with astronomical observations, we have to assume some degree of ergodicity, which just says that ensemble averages become spatial averages as the volume becomes infinitely large. This can only hold true if the volume we are averaging over is infinite, which is clearly not true in our case, where we are limited by the speed of light to only see an "observable" Universe. Basically, we are acting as if different parts of the Universe are independent realizations of the underlying random process. For this reason the validity of the ergodic hypothesis depends on the ratio of the scale over which we perform the spatial average to the scale at which spatial correlations become negligibly small. When the surveyed volume contains many statistically independent subsamples, ergodicity is expected to hold. This, however, will introduce statistical errors called sample variance.
Let's summarise how we can actually compare theory with experiment. On the observational side of things, we divide the sky into patches, calculate the product $\delta_m(\mathbf{x})\delta_m(\mathbf{x}')$, then average it over a number of patches given by the ratio of the size of the observable Universe to the patch size. The result is supposed to approximate

(8.48), with an error that's bigger on larger scales (since the patches are bigger, i.e. they are less in number) rather than on small scales. Once we have a $\xi_{\rm obs}(r)$, we Fourier transform it to the power spectrum (8.46), which is finally compared to its theoretical prediction. At present time, we seem to have a rather strong agreement between theory and observations, as is seen in Figure 8.1.

We have said above how the late time matter power spectrum is a result of the evolution of a primordial power spectrum. In general, scalar perturbations generated during inflation are commonly parametrized in terms of $P_{\mathcal{R}}$, since the gauge-invariant quantity $\mathcal{R}$ has the advantage of being conserved on super-horizon scales, and is thus an unambiguous anchoring point. To introduce standard nomenclature, consider that from the relation (8.22), it follows that:

$$P_{\mathcal{R}}(k) = \frac{9}{4} P_{\Psi} \tag{8.51}$$

Using the result from (8.24) and writing it in terms of the slow-roll variable ϵ as defined in (5.54), we have:

$$P_{\mathcal{R}}(k) = \frac{2\pi G}{k^3} \frac{H^2}{\epsilon}\Big|_{aH=k} \equiv 2\pi^2 \mathcal{A}_s k^{-3} \left(\frac{k}{k_p}\right)^{n_s - 1} \tag{8.52}$$

where $\mathcal{A}_s$ is the amplitude of the curvature perturbations in a logarithmic wavenumber interval centered around the pivot scale k_p, and n_s is the tilt (or spectral index). Notice that these two parameters are part of the standard ΛCDM parameters as introduced in (1.5). The pivot scale k_p is a matter of convention, and is ofted chosen to be a scale most constrained by observations (for instance, for CMB anisotropies, the Planck team adopts $k_p = 0.05\ \mathrm{Mpc}^{-1}$). Observations found that:

$$\mathcal{A}_s = \frac{k_p^3}{2\pi^2} P_{\mathcal{R}}(k_p) \simeq 10^{-9} \tag{8.53}$$

Thus, the typical amplitude of scalar perturbations on the scale k_p is $\sqrt{\mathcal{A}_s} \sim 10^{-4}$, which is the order of magnitude mentioned above.

The spectral index $n_s = 1$ is particularly interesting, since it gives rise to what is called a "scale invariant" power spectrum. Consider, in fact:

$$\langle \Phi(\lambda \mathbf{x}) \Phi(\lambda \mathbf{x}') \rangle = \int \frac{d^3k}{(2\pi)^3} \underbrace{\frac{1}{k^3}}_{\mathcal{P}_{\Phi}(k, t_i)} e^{-i\lambda \mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} = \langle \Phi(\mathbf{x}) \Phi(\mathbf{x}') \rangle \tag{8.54}$$

This means that the correlation between any two points is independent of their distance, i.e. there are correlations at all scales. Notice that this is different from white noise, where there is power at all scales, so no correlation at any scale.

One final note about the matter power spectrum regards the so-called "dimensionless power spectrum". In the limit $r \to 0$, the two point function reduces to the variance (which is what we used in (8.49)):

$$\sigma^2 = \langle \delta(\mathbf{x})^2 \rangle \tag{8.55}$$

Since the power spectrum and the correlation function are Fourier-related, it is expected that a further interpretation of the power spectrum has to do with its contribution to the variance. In fact, we can also use the limit in equation (8.47):

$$\sigma^2 = \frac{1}{2\pi^2}\int_0^\infty k^2 P_m(k) dk \tag{8.56}$$

This equation is telling us that the product $P_m(k)/2\pi^2$ gives the power, i.e. the contribution to the variance, per unit k-space volume due to the modes with wave-number k. However, a thin spherical shell in k-space contains many different modes and has a k-space volume $4\pi k^2 dk$, so that the total power contributed by perturbations with wave-number between k and $k + dk$ is proportional to the power spectrum and the number (volume) of contributing modes. In general it is convenient to introduce a new quantity, the dimensionless power spectrum:

$$\Delta^2(k) = \frac{k^3}{2\pi^2} P_m(k) \tag{8.57}$$

which is the contribution to the variance per bin of $\log k$, instead of unit k-space. Therefore, for instance, $\Delta^2(k) = 1$ is just shorthand for saying that the Fourier modes in a unit logarithmic bin around wave-number k generate fluctuations δ_m of order unity. Summarizing: small values of Δ^2 correspond to small density contrasts, while $\Delta^2 \sim 1$ indicates an over-density comparable to the average. A quick manipulation of the $P(k)$ data in 8.1 shows that Δ^2 rises to large values in the large k region. In other words, the data show that the Universe exhibits large inhomogeneities on small scales (large k).
Notice a final subtle point. In the definition of the inflaton's power spectrum (8.17), we have introduced $P_{\delta\phi}$ through a vacuum expectation value. On the other hand, as we have just argued, the power spectra are defined via ensemble averages. These two are not the same. However, we can equate them if our Universe is only a small part of a larger patch of spacetime that went through an era of inflation.

8.4 Linear Evolution: an Overview

The evolution of perturbations can be divided into roughly three separate stages.
Early on, right after inflation, all modes of interest are outside the horizon and the perturbations are all frozen, i.e. they cannot grow, because of causality reasons, and are thus constant in time. At intermediate times, two things happen: some scales enter the horizon and the Universe transitions from an era of radiation domination ($a \ll a_{\rm eq}$) to an era of matter domination ($a \gg a_{\rm eq}$). It is important to understand that the evolution of a mode entering the horizon during the matter era differs significantly from that of a mode crossing the horizon during the radiation era. Finally, deep in the matter era, all modes evolve identically once again, remaining constant before decaying once dark energy becomes the dominating energy density. In Figure 8.2, the evolution of the potential Φ is plotted, such that these stages are clearly shown.

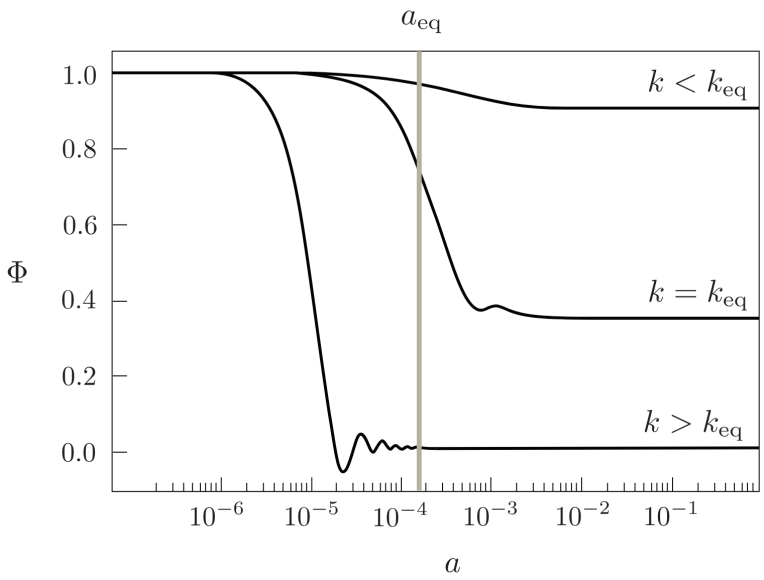

FIGURE 8.2: Linear evolution of the gravitational potential for three different wave-modes.

As far as the power spectrum is concerned, we are only able to observe the distribution of matter mostly at late epochs, in the third stage of evolution. The goal is to have an expression for the late-time matter power spectrum as a function of the initial power spectrum sourced by inflation, so it is convenient to start with the potential Φ, which is schematically written as:

$$\Phi(\mathbf{k}, a) = \frac{3}{5}\mathcal{R}(\mathbf{k}) \times [\text{Transfer function}(k)] \times [\text{Growth factor}(a)] \tag{8.58}$$

where the factor of $3/5$ will be derived below. Note that this is a definition of the "transfer function" and the "growth factor". The transfer function describes the amount the potential decreases by once it reaches an era deep into matter domination. In fact, since even the largest wavelengths perturbations decrease slightly from their initial value, as Figure 8.2 shows, the transfer function is defined so that on large-scales it equals one:

$$T(k) \equiv \frac{\Phi(\mathbf{k}, a)}{\Phi_{\text{large-scale}}(\mathbf{k}, a_{\text{late}})} \tag{8.59}$$

where a_{late} is, again, some time during the third stage of evolution (it can be, say, $a_{\text{late}} \sim 10^{-2} - 10^{-1}$), and $\Phi_{\text{large-scale}}$ is technically the solution for the potential for modes that crossed the horizon deep into the matter epoch. The growth factor is instead defined by the ratio of the potential to its value right after the transfer function regime:

$$\frac{\Phi(\mathbf{k}, a)}{\Phi(\mathbf{k}, a_{\text{late}})} \equiv \frac{D_+(a)}{a} \quad (a > a_{\text{late}}) \tag{8.60}$$

where D_+ is the "growth factor". Using these conventions, we have that:

$$\Phi(\mathbf{k}, a) = \frac{3}{5}\mathcal{R}(\mathbf{k})T(k)\frac{D_+(a)}{a} \quad (a > a_{\text{late}}) \tag{8.61}$$

The evolution of CDM strictly follows from the evolution of the potential via Einstein's equations. Baryons follow closely the dark matter perturbations, so instead of only using the CDM component, during late-times it is customary to group the two together under δ_m. A combination of Einstein's equations (8.2) results in an algebraic equation for the potential that resembles Poisson's equation:

$$k^2\Phi = 4\pi G a^2 \left[\rho_m \delta_m + 4\rho_r \Theta_{r,0} + \frac{3aH}{k}(i\rho_m u_m + 4\rho_r \Theta_{r,1})\right] \tag{8.62}$$

The large-k, no radiation limit of (8.62) gives:

$$k^2\Phi = 4\pi G \rho_m a^2 \delta_m \tag{8.63}$$

While this equation is no longer correct for k of order aH or less, for large-scale structure applications this is a valid limit, as the most precise measurements are for modes that satisfy $k \gg aH$. Now, the matter density satisfies $\rho_m = \Omega_m \rho_{\rm cr}/a^3$ and, by definition of the critical density, $4\pi G\rho_{\rm cr} = (3/2)H_0^2$. Therefore:

$$\delta_m(\mathbf{k}, a) = \frac{2k^2 a}{3\Omega_m H_0^2}\Phi(\mathbf{k}, a) \quad (a > a_{\rm late}, k \gg aH) \tag{8.64}$$

This equation justifies the name "growth" factor: while the potential stays constant, the matter density contrast δ_m grows. However, note that this means that the density fluctuation $\delta\rho_m$ is still decreasing, since $\delta\rho_m = \bar{\rho}_m \delta_m \propto a^{-2}$, but not as fast as the background density, $\bar{\rho}_m \propto a^{-3}$. Using the functional form of the potential given in (8.61), the density contrast (8.64) becomes:

$$\delta_m(\mathbf{k}, a) = \frac{2}{5}\frac{k^2}{\Omega_m H_0^2}\mathcal{R}(\mathbf{k})T(k)D_+(a) \quad (a > a_{\rm late}, k \gg aH) \tag{8.65}$$

The definition of the matter power spectrum (8.1) and the fact that $\mathcal{R}(\mathbf{k})$ is drawn from a Gaussian distribution with power spectrum (8.52) implies that:

$$P_m(k, a) = \frac{8\pi^2}{25}\frac{\mathcal{A}_s}{\Omega_m^2} D_+^2(a) T^2(k) \frac{k^{n_s}}{H_0^4} k_p^{n_s - 1} \quad (a > a_{\rm late}, k \gg aH) \tag{8.66}$$

The goal of this chapter is therefore to derive expressions for the growth factor and the transfer function. The end result will be the black line in Figure 8.1.

Let us now consider the equations necessary to follow the evolution of dark matter perturbations. First, as suggested above, before recombination ($a < a_\star$) the photon and neutrino energy perturbations will only be considered through their monopole and dipole moments, while all higher moments will be neglected. This is a valid approximation for the photons, while neutrinos free stream and are not tightly coupled. Still, neglecting higher neutrino moments is better than neglecting neutrinos altogether. Also, neutrino masses will be set to zero. After recombination ($a > a_\star$), radiation can be safely neglected altogether, since CDM interacts with other species only through gravity, and the potential Φ is only sourced by matter densities (see, for instance, (8.68) in the no-radiation limit).

Neutrinos and photons, if they are only characterized by their monopoles and dipoles, obey the same equation and, if their distributions start out with the same adiabatic

initial conditions, we can combine them and proceed with the total monopole $\Theta_{0,r}$ and total dipole $\Theta_{1,r}$ moments. The desired set of Boltzmann equations are therefore:

$$\begin{aligned} \Theta'_{r,0} + k\Theta_{r,1} &= -\Phi' \\ \Theta'_{r,1} - \frac{k}{3}\Theta_{r,0} &= -\frac{k}{3}\Phi \\ \delta'_c + iku_c &= -3\Phi' \\ u'_c + \mathcal{H}u_c &= ik\Phi \end{aligned} \tag{8.67}$$

Obtaining the first two equations from (8.4) requires some work. In practice, the tight-coupling approximation was used to get rid of baryon densities, which is a fairly good approximation before recombination. The effect of baryons (together with neutrino masses) will be explored at the end of this chapter. Also, since there are no quadrupole moments, the second Einstein equation implies $\Phi = -\Psi$, as we have argued up until now. The evolution of the potential is instead given by the first Einstein equation:

$$k^2\Phi + 3\mathcal{H}(\Phi' + \mathcal{H}\Phi) = 4\pi G a^2[\rho_c\delta_c + 4\rho_r\Theta_{r,0}] \tag{8.68}$$

or the alternative algebraic equation (8.62):

$$k^2\Phi = 4\pi G a^2\left[\rho_c\delta_c + 4\rho_r\Theta_{r,0} + \frac{3aH}{k}(i\rho_c u_c + 4\rho_r\Theta_{r,1})\right] \tag{8.69}$$

While the most obvious way of solving these equations would be numerically, analytical partial solutions offer more insights, despite being harder to come by. In particular, an analytical solution valid for all scales at all times is not available. To make progress, simplifying assumptions will be made to reduce the total number of equations, which will be valid only for a specific time, or specific scales. In the end, all the approximate solutions will be patched up together to form a unique solution valid everywhere in the parameter space. Regimes where useful physical approximations can be made are:

- Super-horizon regime: when $k\eta \ll 1$, analytical solutions are known for all times, even through matter-radiation equality;
- Horizon crossing: a mode crossing the horizon can be followed only for very early or very late times, that is when the equation of state is constant;
- Sub-horizon regime: when $k\eta \gg 1$, analytical solutions are known for all times, even through matter-radiation equality;

The evolution at late and early times is therefore known analytically, combining all three regimes together. Due to lack of simplifying limits, the only missing evolution is for those intermediate modes that cross the horizon during matter-radiation equality.

8.5 Large Scales

In this section we analyze the evolution of scales that cross the horizon deep into the matter era (indeed, this is what we mean by the loosely defined term "large" scales).

First, we study the super-horizon solution, which is available analytically, as we said before. Then, using this same solution, following the mode as it enters the horizon will be straightforward.

Super-horizon evolution

The regime of interest is now $k\eta \ll 1$, which corresponds to the limit taken for the initial conditions. Indeed, the velocities and monopoles decouple from the evolution, as should be clear by looking at the Boltzmann equations (8.67). Since the algebraic Einstein equation contains terms proportional to k^{-1}, we use instead its counterpart (8.68). We are then left with the following set of coupled equations:

$$\begin{aligned} \Theta'_{r,0} &= -\Phi \\ \delta'_c &= -3\Phi' \\ 3\mathcal{H}(\Phi' + \mathcal{H}\Phi) &= 4\pi G a^2[\rho_c\delta_c + 4\rho_r\Theta_{r,0}] \end{aligned} \tag{8.70}$$

The first two equations require the quantity $\delta_c - 3\Theta_{0,r}$ to be a constant. This constant is, as we argued for the initial conditions, set to zero by the condition of adiabatic perturbations. The Einstein equation can be rewritten using $\Theta_{r,0} = \delta_c/3$ and introducing the quantity $y \equiv a/a_{\rm eq} = \rho_m/\rho_r$:

$$3\mathcal{H}(\Phi' + \mathcal{H}\Phi) = 4\pi G a^2 \rho_c \delta_c \left[1 + \frac{4}{3y}\right] \tag{8.71}$$

The two first-order differential equations for Φ and δ_c can be turned into one second-order equation for the potential. This is done in the following way. In terms of y and derivatives with respect to it, (8.71) becomes:

$$y\frac{d\Phi}{dy} + \Phi = \frac{3y+4}{6(y+1)}\delta_c \tag{8.72}$$

where we have used $d/d\eta = aHy\, d/dy$ and that $8\pi G\rho_c/3 = H^2 y/(y+1)$. Now, we turn (8.72) into an expression for δ_c and then differentiate it with respect to y. Then, the dark matter equation sets $d\delta_c/dy$ equal to $-3d\Phi/dy$. This leads to:

$$\frac{d^2\Phi}{dy} = \frac{21y^2 + 54y + 32}{2y(y+1)(3y+4)}\frac{d\Phi}{dy} + \frac{\Phi}{y(y+1)(3y+4)} = 0 \tag{8.73}$$

This equation admits an analytic solution, which takes the form:

$$\Phi(\mathbf{k}, y) = \frac{1}{10y^3}[16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16]\Phi(\mathbf{k}, 0) \tag{8.74}$$

The form of (8.74) is the final expression for the potential on super-horizon scales. At small y, i.e. early times, this expression fixes $\Phi \simeq \Phi(0)$, consistent with the potential being constant on the far left in Figure 8.2. At late times, once the Universe has become matter dominated, the limit for large y of (8.74) becomes $\Phi \simeq (9/10)\Phi(0)$. This $9/10$ term is the factor by which the potential drops after matter-radiation equality that was mentioned above. This result allows us to obtain useful relations between the super-horizon gravitational potential and the curvature perturbation $\mathcal{R}$. Above,

we found that, in the radiation dominated era, $2/3\mathcal{R} = -\Psi = \Phi$. Therefore, in the matter era the relation becomes $\Phi = (9/10)(2/3)\mathcal{R}$, so that:

$$\Phi(\mathbf{k},\eta)\Big|_{\text{Super-horizon}} = \begin{cases} \frac{2}{3}\mathcal{R}(\mathbf{k}) & \text{Radiation domination} \\ \frac{3}{5}\mathcal{R}(\mathbf{k}) & \text{Matter domination} \end{cases} \tag{8.75}$$

Below, we will find that, in the matter era, the potential does not evolve even when inside the horizon. This explains the factor of $3/5$ in the definition (8.58).

Horizon crossing

In this section we would like to show that the potential remains constant as the (large-scale) modes enter the horizon. This holds true if we consider a period deep into matter domination, when $a \gg a_{\text{eq}}$.
The key equations are those in (8.67), in the limit where radiation is negligible (see discussion above), so that the two radiation equations will be neglected. The two matter equations are then supplemented with the Einstein equation, this time in the algebraic form (8.69). Since this equation is algebraic, it can be used to replace Φ in the matter equation, so that we are left with a coupled system of differential equations in δ_c and u_c. Instead of solving these equations directly, we use a trick to find that the potential remains constant during horizon crossing. Since super horizon modes are constant, the initial condition for our problem is $\Phi(0)' = 0$. Therefore, if one is able to show that $\Phi = \text{const}$ is a solution of the two differential equations for matter, it will also be the full solution far in the matter era (this is true as long as other solutions to the same equations are decaying, which they are). Summarizing, we need to prove that the following set of equations:

$$\begin{aligned} \delta_c' + iku_c &= 0 \\ u_c' + aHu_c &= ik\Phi \\ k^2\Phi &= \frac{3}{2}a^2H^2\left[\delta_c + \frac{3aHiu_c}{k}\right] \end{aligned} \tag{8.76}$$

admits a solution with $\Phi' = 0$. Replacing δ_c in the first of these equations with Φ and u_c, taken from the last, and using $H^2 \propto a^{-3/2}$ for the matter era, one finds:

$$\frac{2k^2\Phi'}{3a^2H^2} + \frac{2k^2\Phi}{3aH} - \frac{3aHiu_c'}{k} + \frac{3a^2H^2iu_c}{2k} + iku_c = 0 \tag{8.77}$$

This equation can be turned into a second-order differential equation in Φ by using the equation governing u_c. Eliminating u_c' in (8.77) yields:

$$\frac{2k^2\Phi'}{3a^2H^2} + \left[\frac{iu_c}{k} + \frac{2\Phi}{3aH}\right]\left[\frac{9a^2H^2}{2} + k^2\right] = 0 \tag{8.78}$$

Clearly, if the second-order equation does not contain terms proportional to Φ, that is it takes the form $\alpha\Phi'' + \beta\Phi' = 0$, then $\Phi' = 0$ will actually be a valid solution. Differentiating (8.78) with respect to η and only keeping terms proportional to Φ

(neglecting all its derivatives), one finds:

$$\left[\frac{iu_c'}{k}+\frac{\Phi}{3}\right]\left[\frac{9a^2H^2}{2}+k^2\right]+\left[\frac{iu_c}{k}+\frac{2\Phi}{3aH}\right]\frac{d}{d\eta}\frac{9a^2H^2}{2} = -\left[\frac{iaHu_c}{k}+\frac{2\Phi}{3}\right](9a^2H^2+k^2) \tag{8.79}$$

where u_c' was eliminated in favor of u_c using the CDM velocity equation again. But (8.78) implies that the term in brackets in the above equation is actually proportional to Φ', so that the second-order equation for Φ does not contain any terms proportional to Φ. Therefore, the solution $\Phi = \text{const}$ is the solution in the matter era.
Potential wells in the Universe, in the matter era, don't deepen (grow) as time moves on. This is only possible if the two effects, namely gravity and the expansion of the Universe, precisely balance their effects, such that matter accretes just fast enough to balance the disrupting stretch of space. In contrast, from Figure 8.2, it is clear that potential wells in the radiation era (when the mode enters the horizon) decrease drastically, as the expansion of the Universe is too fast for matter to be able to cluster. This is also reflected in the fact that matter perturbations only increase logarithmically, $\delta_m(a) \propto \log a$, in the radiation era (see below), rather than linearly, $\delta_m(a) \propto D_+(a) \propto a$, in the matter era. The results of this discussion implies that the transfer function is close to unity on large-scales, which enter the horizon only in the matter era, that is those modes that satisfy $k < k_{\rm eq} = a_{\rm eq}H(a_{\rm eq})$. Neglecting baryons and anisotropic stresses related to the radiation components, means fundamentally that the transfer function only depends of $k/k_{\rm eq}$.

8.6 Small Scales

In this section we would like to analyse the evolution of perturbations that cross the horizon in the radiation era (by definition, we call these "small"-scales), all the way to present time. The problem splits naturally into two steps, which are the reverse of those for large-scales. Namely, for large scales, we evolved the modes outside the horizon, and only when deep in the matter era did we follow their plunge into the horizon. Here, on the other hand, we first evolve the modes through the horizon, deep in the radiation era, and only later consider their sub-horizon growth. The latter solution will be specified by taking initial conditions from the horizon-crossing solution, just as the large-scale horizon-crossing evolution took as initial condition the value of perturbations outside the horizon.

Horizon crossing

When the Universe is radiation dominated, the potential is sourced by perturbations in the radiation. The evolution of dark matter is then determined by the potential, but it doesn't itself influence the potential. Therefore, first, we solve the coupled equations for the radiation and the potential, and then determine the evolution of CDM. These equations are the first two in (8.67) and the algebraic Einstein equation, which takes the form:

$$\Phi = \frac{6a^2H^2}{k^2}\left[\Theta_{r,0}+\frac{3aH}{k}\Theta_{r,1}\right] \tag{8.80}$$

where we have used that, in the radiation era, $H^2 = 8\pi G\rho_r/3$. Also, recall that in the radiation era, $aH = 1/\eta$. Using (8.80), we can eliminate the monopole from the radiation equations and write them as:

$$\begin{aligned} -\frac{3}{k\eta}\Theta'_{r,1} + k\Theta_{r,1}\left[1+\frac{3}{k^2\eta^2}\right] &= -\Phi\left[1+\frac{k^2\eta^2}{6}\right] - \Phi\frac{k^2\eta}{3} \\ \Theta'_{r,1} + \frac{1}{\eta}\Theta_{r,1} &= -\frac{k}{3}\Phi\left[1-\frac{k^2\eta^2}{6}\right] \end{aligned} \tag{8.81}$$

Combining the two equations above yields a second-order equation for the potential, which reads:

$$\Phi'' + \frac{4}{\eta}\Phi' + \frac{k^2}{3}\Phi = 0 \tag{8.82}$$

The solution is given in terms of the spherical Bessel function of order 1, which can be expressed in terms of trigonometric functions:

$$\Phi(\mathbf{k},\eta) = 2\left(\frac{\sin x - x\cos x}{x^3}\right)_{x=k\eta/\sqrt{3}} \mathcal{R}(\mathbf{k}) \tag{8.83}$$

Where the factor of 2 and the curvature perturbation arise due to the matching to initial conditions $\Phi = 2\mathcal{R}/3$. This solution shows that, as soon as the mode enters the horizon, the potential begins to oscillate and decay, as depicted in Figure 8.2, on the left. This behavior agrees with the qualitative arguments carried above, that alluded to the ability of pressure to counteract gravity. The pressure prevents overdensities from growing, which is already evident from (8.80) if one neglects the monopole well within the horizon, since it reads $\Phi \sim \Theta_0/\eta^2$. Because Θ_0 oscillates with fixed amplitude, Φ oscillates but with an amplitude decreasing as η^{-2}.
Once the form of the potential is known, we can easily understand how the matter perturbations evolve. In fact, turning the two matter equations in (8.67) into one second-order equation with the potential serving as an external source, we get:

$$\delta_c'' + \frac{1}{\eta}\delta_c' = S(k,\eta) \tag{8.84}$$

where:

$$S(k,\eta) = -3\Phi'' + k^2\Phi - \frac{3}{\eta}\Phi' \tag{8.85}$$

is the source term. The two solutions to the homogeneous equation ($S = 0$) are $\delta_c =$ const and $\delta_c = \log a$, which is the logarithmic growth mentioned before. The general solution is a linear combination of a particular solution and the homogeneous solution. The particular solution can be constructed from the two homogeneous solutions (call them s_1 and s_2) and the source term:

$$\delta_c(k,\eta) = C_1 + C_2\log(k\eta) - \int_0^\eta d\tilde{\eta}\; S(k,\tilde{\eta})\tilde{\eta}[\log(k\tilde{\eta}) - \log(k\eta)] \tag{8.86}$$

where the factors of k in the logarithms are for convenience. Initial conditions (δ_c constant), dictate that the coefficient C_2 vanishes and $C_1 = \delta_c(k,\eta=0) = \mathcal{R}$, since the integral is obviously small. As far as the integral is concerned, the integrand

contributes the most when $k\eta \sim 1$, since when the potential is well inside the horizon, it has decayed. Therefore, the integral over $S(\tilde{\eta})\log(k\tilde{\eta})$, as a function of η, asymptotes to a constant, while the integral over $S(\tilde{\eta})\log(k\eta)$ leads lead to a term proportional to $\log(k\eta)$ multiplied by some constant. Therefore, one expects that when the potential has entered the horizon:

$$\delta_c(k,\eta) = A\mathcal{R}\log(Bk\eta) \tag{8.87}$$

The constants A and B can be fixed by referring to the relevant parts of the equations above. In particular, due to the considerations just carried out, we can just solve the following system:

$$\begin{aligned} A\mathcal{R}\log(B) &= \mathcal{R} - \int_0^\infty d\eta S(k,\tilde{\eta})\tilde{\eta}\log(k\tilde{\eta}) \\ A\mathcal{R} &= \int_0^\infty d\tilde{\eta} S(k,\tilde{\eta})\tilde{\eta} \end{aligned} \tag{8.88}$$

The upper limit of the integral was set to ∞ in accord with the expectation that the source term vanishes when the potential enters the horizon and decays. The constants were found to be well fitted by $A = 6.4$ and $B = 0.44$.
This result is consistent with the discussions above. The dark matter perturbations in the radiation era grow, but only logarithmically, due to the rapid expansion of the Universe. In contrast, tightly coupled photon and baryon perturbations oscillate up until recombination due to high pressures, and therefore cannot grow.

Sub-horizon evolution

Since radiation perturbations are suppressed once the small-scale modes enter the horizon early in the history of the Universe, eventually the growth in the CDM component offsets the higher mean radiation energy density. That is, despite the radiation still being the main contributor to the densities, the factor $\rho_c\delta_c$ becomes larger than $\rho_r\Theta_{r,0}$. Once this threshold is crossed, the potential and the dark matter evolve according to coupled differential equations, which are not affected by the radiation component. Therefore, one can solve these equations in the limit of $k\eta \gg 1$, that is for modes inside the horizon. Once the sub-horizon evolution is known, it is patched together with the logarithmic growth at early times found above.
In terms of the variable $y = a/a_{\rm eq}$, the two dark matter equations and the algebraic Einstein equation become:

$$\begin{aligned} \frac{d\delta_c}{dy} + \frac{iku_c}{aHy} &= -3\frac{d\Phi}{dy} \\ \frac{du_c}{dy} + \frac{u_c}{y} &= \frac{ik\Phi}{aHy} \\ k^2\Phi &= \frac{3y}{2(y+1)}a^2H^2\delta_c \end{aligned} \tag{8.89}$$

Notice that we neglect here both baryons and dark energy. While the latter is a good approximation for the cosmological times that are being considered, baryons do induce differences from the correct numerical result. As usual, the first two equations

can be combined into one second-order equation in the CDM density contrast:

$$\frac{d^2\delta_c}{dy^2} - \frac{ik(2+3y)u_c}{2aHy^2(1+y)} = -3\frac{d^2\Phi}{dy^2} + \frac{k^2\Phi}{a^2H^2y^2} \tag{8.90}$$

In this derivation we have used the fact that $d(1/aHy)/dy = -(1+y)^{-1}(2aHy)^{-1}$. The first term on the right is negligible compared to the second one, given that the latter is multiplied by $k/aH \gg 1$. Then, using the Einstein equation, we recognize this second term on the right hand side as $3\delta_c/[2y(y+1)]$. Finally, the velocity in the second term on the left hand side can be eliminated via the density contrast equation, but neglecting the potential, since it is small compared to δ_c when inside the horizon. Then (8.90) becomes:

$$\frac{d^2\delta_c}{dy^2} + \frac{2+3y}{2y(y+1)}\frac{d\delta_c}{dy} - \frac{3}{2y(y+1)}\delta_c = 0 \tag{8.91}$$

This equation is known as the "Meszaros equation". The two independent solutions to this equation need to be matched on to the logarithmic mode established above. To find the solutions, we can use our knowledge of the evolution of δ_c in the matter era, inside the horizon. Above, these have been shown to grow linearly with the scale factor, so one of the solutions to (8.91) must be a first-order polynomial in the variable y. For this mode, the second derivative of δ_c vanishes, so that its governing equation is:

$$\frac{1}{\delta_{c,+}}\frac{d\delta_{c,+}}{dy} = \frac{3}{2+3y} \tag{8.92}$$

the solution of which is $\delta_{c,+} \propto y + 2/3$, or:

$$D_+(a) = a + \frac{2}{3}a_{\rm eq} \tag{8.93}$$

Since this solution reduces to $D_+ \simeq a$ for $y \gg 1$, we have identified it with the growth factor introduced in (8.60). Notice that, since we are ignoring dark energy, this form for the growth factor can only be valid up until dark energy comes to dominate the Universe, that is $a \leq 0.1$. The second solution to the Meszaros equation is obtained easily by a change of variables, and reads:

$$D_-(y) = (y + \frac{2}{3})\log\left[\frac{\sqrt{y-1}+1}{\sqrt{1+y}-1}\right] - 2\sqrt{1+y} \tag{8.94}$$

At early times ($y \ll 1$), $D_+ = $ const, while D_- is proportional to $\log y$, as expected. At late times, ($y \gg 1$), the growing solution grows linearly, whereas the decaying mode falls off as $y^{-3/2}$. The general solution is a linear combination of D_+ and D_-:

$$\delta_c(k,y) = C_1 D_+(y) + C_2 D_-(y) \quad y \gg y_H \tag{8.95}$$

where $y_H = a_H/a_{\rm eq}$ is the scale factor when the mode enters the horizon, divided by $a_{\rm eq}$. The constants C_1 and C_2 are determined by a matching with the logarithmic solution at early times, which is only valid for those modes that enter the horizon well before equality $y_H \ll y \ll 1$, so the full sub-horizon solution can only be found for those modes. For modes that cross the horizon near equality, no analytical solution

is known, as stated above. Anyway, the matching between the two solutions and their derivatives read:

$$\begin{aligned} A\mathcal{R}\log(By_m/y_H) &= C_1 D_+(y_m) + C_2 D_-(y_m) \\ \frac{A\mathcal{R}}{y_m} &= C_1 D'_+(y_m) + C_2 D'_-(y_m) \end{aligned} \tag{8.96}$$

where, again, the matching epoch y_m must satisfy $y_H \ll y_m \ll 1$. The argument of the logarithm was replaced by y/y_H, which is only valid as long as the matching epoch is deep in the radiation era.
Numerical simulations for the various components in the Universe find the following evolution for the different perturbations:

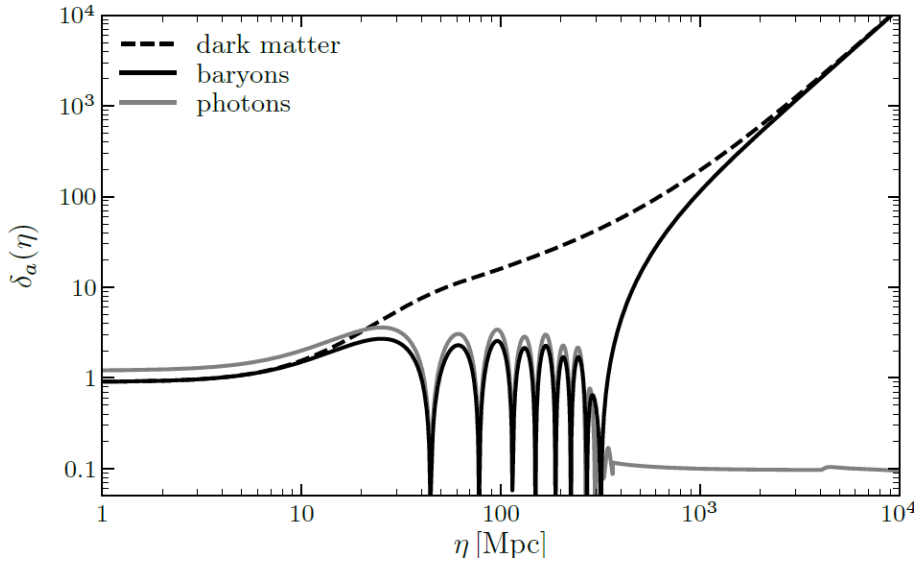

FIGURE 8.3: Evolution of perturbations for the various species in the Universe..

This Figure represents all the results obtained in this chapter. The CDM perturbations grow early on, first logarithmically, then linearly, due to dark matter being pressureless and non-interacting. Photons and baryons, on the other hand, are tightly coupled before recombination and therefore oscillate. Photons, once decoupled, free stream in the Universe, while baryons quickly match the CDM perturbations.

8.7 Numerical Results and Fits

The goal of this chapter was to find solutions for the transfer function and the growth factor. Using all the results obtained until now, it is possible to find expressions and fits for these functions. While the growth factor was already found in (8.93), we would like to analyse corrections to this equation due to dark energy at late times.

The transfer function

First, we transform (8.95) in an expression for the transfer function. Since this is defined by the behavior of matter only at late times, in reality we only care about

the growing mode, while the decaying mode will have vanished by then. In other words, the important constant in this case is C_1, which is found via the system of equations (8.96):

$$C_1 = \frac{D'_-(y_m)\log(By_m/y_H) - D_-(y_m)/y_m}{D_+(y_m)D'_-(y_m) - D'_+(y_m)D_-(y_m)} A\mathcal{R} \tag{8.97}$$

The denominator, using the expressions for D_+ and D_- in (8.93) and (8.94) respectively, reduces to $-4/9y_m$ in the limit $y_m \ll 1$, the matching epoch. Similarly, the same limit implies that $D_- \to (2/3)\log(4/y) - 2$ and $D_- \to -2/3y$, so that C_1 becomes:

$$C_1 \to -\frac{9}{4}A\mathcal{R}\left[-\frac{2}{3}\log(\frac{By_m}{y_H}) - \frac{2}{3}\log(\frac{4}{y_m}) + 2\right] \tag{8.98}$$

At late times, then, the density contrast for dark matter grows as:

$$\delta_c(\mathbf{k}, a) = \frac{3}{2}A\mathcal{R}(\mathbf{k})\log\left[\frac{4Be^{-3}a_{\rm eq}}{a_H}\right] D_+(a) \quad a \gg a_{\rm eq} \tag{8.99}$$

Since, in the limit of no-baryons we have that $\delta_c = \delta_m$, comparing the above equation with (8.65) immediately yields the following form for the transfer function:

$$T(k) = \frac{15}{4}\frac{\Omega_m H_0^2}{k^2 a_{\rm eq}} A\log\left[\frac{4Be^{-3}\sqrt{2}k}{k_{\rm eq}}\right] \quad k \gg k_{\rm eq} \tag{8.100}$$

Plugging in numbers ($A = 6.4$ and $B = 0.44$), and using that $k_{\rm eq} = \sqrt{2\Omega_m}H_0 a_{\rm eq}^{-1/2}$:

$$T(k) = 12\frac{k_{\rm eq^2}}{k^2}\log\left[0.12\frac{k}{k_{\rm eq}}\right] \tag{8.101}$$

As expected, the transfer function, when baryons are neglected, only depends on the ratio $k/k_{\rm eq}$. More sophisticated fitting techniques, like the BBKS fit, find the following form for the transfer function:

$$T(k) = \frac{\log(1+2.34q)}{2.34q}[1 + 3.89q + (16.1q)^2 + (5.46q)^3 + (6.71q)^4]^{1/4} \tag{8.102}$$

where:

$$q \equiv \frac{1}{\Gamma}\left(\frac{k}{h\ {\rm Mpc}^{-1}}\right) \quad \Gamma = h\Omega_m \tag{8.103}$$

A realistic CDM model obviously also needs to include baryons, the presence of which influences the transfer function. Increasing the baryonic mass fraction largely leaves the shape of T(k) intact, but it causes a reduction of the shape parameter, which is well approximated by:

$$\Gamma = \Omega_m h \exp\left[-\Omega_b(1 + \sqrt{2h}/\Omega_m)\right] \tag{8.104}$$

However, if the baryonic mass fraction becomes sufficiently large, the transfer function starts to develop oscillations, and the BBKS fit is no longer appropriate. In this

case one has to resort to more sophisticated methods. The "baryon acoustic oscillations" (BAO) in the transfer function produce oscillatory features in the matter power spectrum. Such oscillations have indeed been observed in the galaxy distribution on large scales.
Today, fast numerical simulations can compute the transfer function easily, so fits have become obsolete for the most part. However, analytical methods such as the one presented here offer important insights in terms of the underlying physics at play. In fact, the small-scale behavior of $T(k)$ is only understood if one keeps in mind the logarithmic growth in the radiation era. Had there been no such growth, the modes that entered the horizon very early on would have experienced growth only starting from the era of equality, such that their amplitude, relative to large-scale modes, would have been suppressed by a factor of order $(k_{\rm eq}/k)^2$.
The approximations we have made along the way actually have some consequences for the form of the transfer function. In practice, incorporating anisotropic stresses, such that $\Psi \neq -\Phi$, changes the factor 9/10 by which the potential drops for large-scale modes to roughly 0.86, inducing a rise in the small-scale transfer function. Furthermore, adding baryons leads to even more small-scale changes, as we discussed above, such as BAO.

The growth factor

The second part of structure formation is the scale-independent growth factor. Were it not for dark energy and neutrinos, the Meszaros equation would be a valid method to find the form of D_+. In this section, we include the effect of dark energy but still neglect neutrinos, whose effect will be mentioned later.
First, we take the matter equations in (8.67) in the matter era and in the limit $k\eta \gg 1$, that is for all modes of interest inside the horizon. Furthermore, we can neglect the pressure due to baryons at late times, so that they follow equations that look like those of dark matter. Also, baryons follow closely the dark matter perturbations at late times (see Figure 8.3), so that the matter sector can be described through a total matter perturbation variable, defined as $\rho_m\delta_m = \rho_c\delta_c + \rho_b\delta_b$, and similarly for the matter velocities $u_m = (\rho_c u_c + \rho_b u_b)/\rho_m$. Now, multiply the density contrast equation by a and take the derivative with respect to η. Neglecting the right-hand side on sub-horizon scales, one has:

$$[a\delta_m'(\mathbf{k},\eta)]' = ak^2\Phi(\mathbf{k},\eta) \tag{8.105}$$

The potential satisfies the Einstein equation, which we take to be (8.68), in the matter era and $k \gg aH$ limits (that is, it becomes (8.62)):

$$k^2\Phi(\mathbf{k},\eta) = 4\pi G a^2 \rho_m(\eta)\delta_m(\mathbf{k},\eta) \tag{8.106}$$

Using the scaling $\rho_m \propto a^{-3}$ and the definition for Ω_m, the evolution equation for δ_m reads:

$$[a\delta_m']' = \frac{3}{2}\Omega_m H_0^2 \delta_m \tag{8.107}$$

Changing the variable from η to a and carrying out the derivatives, this equation becomes:

$$\frac{d^2\delta_m}{da^2} + \frac{d\log(a^3H)}{da}\frac{d\delta_m}{da} - \frac{3\Omega_m H_0^2}{2a^5H^2}\delta_m = 0 \tag{8.108}$$

While in general (8.108) needs to be solved numerically, when the only components in the Universe are matter, dark energy and curvature (a good approximation at late times), the analytical solution is:

$$D_+(a) = \frac{5\Omega_m}{2}\frac{H}{H_0}\int_0^a da' \frac{1}{(a'H(a')/H_0)^3} \quad (\Lambda, \Omega_k) \tag{8.109}$$

where the initial conditions were set such that, when matter dominates early on, the growth factor should be proportional to a. Notice that this form is a solution of (8.108) only if dark energy is a cosmological constant. The following plot shows the deviation of the growth factor D_+ from the no-dark energy limit $D_+ \propto a$:

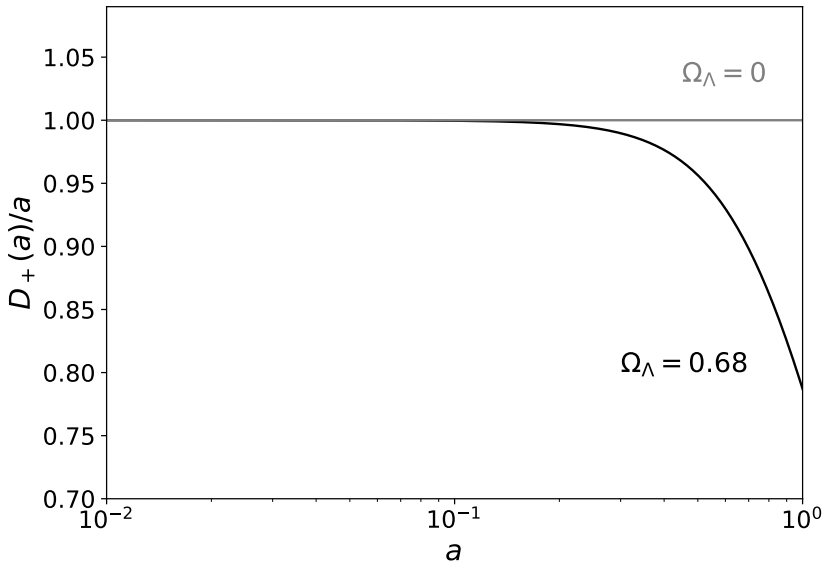

FIGURE 8.4: Growth factor, with and without the dark energy component

8.8 Beyond the Approximations

In this last section we analyze the effect of components that were neglected in this treatment of dark matter perturbations. In particular, we briefly state the effects of baryons, neutrino masses and dark energy.

Baryons

One of the effects of baryons was stated above, and consists in inducing oscillations in the transfer function at the scale of the sound horizon at recombination. These oscillations are called baryon acoustic oscillations and they have been detected in the clustering of galaxies. The oscillations are clearly reflected in the matter power spectrum. The second effect of baryons is a suppression of the transfer function at small scales relative to the no-baryon case. This shouldn't come as a surprise: before decoupling, baryons are tightly coupled to the photons, so that their perturbations are suppressed compared to the dark matter. After recombination, the baryons are released and catch up with the dark matter perturbations (this fact will not be derived here), and fall into the gravitational wells that are the dark matter halos. The

depth of these wells is smaller than estimated above, though, because only a fraction of the total mass of the Universe, Ω_c/Ω_m, was involved in the collapse.

Neutrino masses

The masses of neutrinos affect structure formation in two ways. First, the scaling of their energy density, since they become non-relativistic at late times, switches from a^{-4} to a^{-3}, affecting the evolution of the background through the Friedmann equation. In addition, this fact slightly modifies the growth factor, since it depends on $H(a)$. The second effect is that, since neutrinos move fast, they damp the growth of small-scale structures. Perturbations on scales smaller than the typical distances neutrinos travel, called "free-streaming scales", are therefore suppressed. Furthermore, the growth factor becomes scale-dependent in the presence of massive neutrinos, spoiling the neat decomposition of (8.58).

Dark energy

The main physical effect of dark energy is its impact on the growth factor through the Hubble parameter $H(a)$. Therefore, different models for dark energy predict different growth factors. Parametrizing the dark energy by its equation of state w (which we assume to be constant), the Hubble parameter in a Euclidian Universe reads:

$$\frac{H(z)}{H_0} = \left[\frac{\Omega_m}{a^3} + \frac{\Omega_{\rm DE}}{a^{3(1+w)}}\right] \tag{8.110}$$

at late times. Using this dependence, solving (8.108) numerically leads directly to the desired growth factor. This effect, coupled with other CMB observables, forms the basis of current dark energy constraints arising from large-scale structure.

Part 3: Structure Formation

Chapter 9

Nonlinear Clustering

So far, our treatment has been purely concerned with the linear regime, where $\delta \ll 1$, however, many objects in the present-day Universe, such as galaxies and clusters of galaxies, have densities that are much higher than the average density of the Universe. Tt follows that if we want to understand the formation of these structures, we must go beyond the linear perturbation theory and explore the nonlinear territory, where $\delta \gg 1$.
In this section we thus address the problem of gravitational collapse of overdensities in the nonlinear regime.
As a general rule, the nonlinear gravitational dynamics is difficult to deal with from an analytical standpoint, and as a consequence the last word has to be left to detailed computer $N-$body simulations; the problem with numerical simulations, however, is that they tend to obscure the basic physics contained in the equations, so that they act essentially as a "black box". Luckily, if certain assumptions about the symmetry of the problem are made, simple analytical models can be studied, which give ultimately some insights into the complex physics of gravitational clustering.
In the following we discuss a simple model, the spherical collapse model, which serves both as an example where we can follow analytically the nonlinear evolution, and as a bridge to the so called "Press-Schechter theory", which utilizes the results of the spherical collapse model, and follow up with "excursion set theory", which formalizes Press-Schechter and allows us to be even more precise.

9.1 The Spherical Collapse Model

Consider a flat, matter dominated universe. The average density is equal to the critical density and evolves in times as:

$$\bar{\rho}(t) = \frac{1}{6\pi G}\frac{1}{t^2} \tag{9.1}$$

where we used the results for the Einstein-de Sitter Universe obtained in equation (3.59).
At some initial time t_i we create a spherically symmetric matter overdensity by compressing a region of radius R_i to one with a smaller radius $r_i < R_i$. Using the conservation of mass, we find that the initial density of the perturbation is:

$$\rho_i = \frac{\bar{\rho}_i R_i^3}{r_i^3} \equiv \bar{\rho}_i(1+\delta_i) \tag{9.2}$$

Note that we just introduced the initial density contrast δ_i, but it will not be restricted to values $\delta_i \ll 1$, like it was in the previous sections.
The important point in this problem is that we are requiring spherical symmetry, from which follow some simplifications. First, the evolution of the background is decoupled from that of the perturbation, so that the former is not influenced by the latter, and secondly we can think of the overdensity as composed of infinite thin shells. Our analysis will remain valid unless the shells cross with one another, in which case the mass contained in the shell is no longer constant, making our assumption (9.2) invalid. The shortcomings of this model will become clear below.
Let us now study the evolution of a single mass shell of radius $r(t)$, which evolves according to the Newtonian equation:

$$\frac{d^2r}{dt^2} = -\frac{GM(r)}{r^2} \tag{9.3}$$

where $M(r)$ is the mass enclosed in the shell. If we integrate this equation we get the conservation of energy:

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 - \frac{GM}{r} = E \tag{9.4}$$

where E is the specific energy of the mass shell. Both E and $M(r)$ are constant throughout the collapse. We will study here the case $E < 0$, so that the overdensity acts like a closed universe with positive curvature. The solutions of this equations were given in parametric form in the matter and curvature solution in the first section:

$$\begin{aligned} r(\theta) &= A(1-\cos\theta) \\ t(\theta) &= B(\theta - \sin\theta) \end{aligned} \tag{9.5}$$

with A and B constants to be determined by the initial conditions: take the solutions (9.5) at early times where $\theta \ll 1$ and expand them to lowest order:

$$\begin{aligned} r(\theta) &= A(\theta^2/2 - \theta^4/24) \\ t(\theta) &= B(\theta^3/6 - \theta^5/120) \end{aligned} \tag{9.6}$$

If we insert them into equations (9.3) and (9.4) we get the following relations:

$$A^3 = GMB^2 \qquad A = \frac{GM}{-2E} \tag{9.7}$$

Now, assume that at the initial time $t_i \ll 1$ the radius and the velocity of the mass shell are r_i and v_i, so that the energy is $E = v_i^2/2 - GM/r_i$. If we insert this into the second expression of (9.7) we get:

$$\frac{r_i}{2A} = 1 - \frac{(v_i/H_i r_i)^2}{1+\delta_i} \tag{9.8}$$

where H_i is the initial Hubble constant, and we used the fact that the average initial δ_i is related to M and r_i by:

$$M = \rho_i \frac{4\pi}{3} r_i^3 = \bar{\rho}_i (1+\delta_i) \frac{4\pi}{3} r_i^3 \tag{9.9}$$

where we have used (9.2). Since M is constant, we can write v_i as:

$$v_i = \frac{dr_i}{dt_i} = H_i r_i \left(1 - \frac{1}{3H_i t_i} \frac{\delta_i}{1+\delta_i} \frac{d\log\delta_i}{d\log t_i}\right) \tag{9.10}$$

At sufficiently early times $t_i \ll 1$ and $\delta_i \ll 1$, we have the following scalings:

$$\begin{aligned} H_i t_i &\approx 2/3 \\ \delta_i &\approx t_i^{2/3} \\ \frac{v_i}{H_i r_i} &\approx 1 - \frac{\delta_i}{3} \end{aligned} \tag{9.11}$$

If we insert these into (9.7) and (9.8), we finally get the equations for A and B as a function of initial conditions:

$$A = \frac{3}{10}\frac{r_i}{\delta_i} \quad B = \frac{3}{4}\left(\frac{3}{5}\right)^{3/2} \frac{1}{\delta_i} \tag{9.12}$$

The motion of the mass shell is therefore specified, in a given cosmology, via the initial conditions on the radius r_i of the mass shell and the mean overdensity δ_i.
The mass shell reaches the maximum expansion at $\theta = \pi$ (the maximum of the function $r(\theta)$) and then turns around. This means that the radius and time at maximum expansion are:

$$r_{\max} = 2A = \frac{6}{10}\frac{r_i}{\delta_i} \qquad t_{\max} = B\pi = \frac{3\pi}{4}\left(\frac{3}{5}\right)^{3/2} \frac{1}{\delta_i} \tag{9.13}$$

To get an idea of how the shell evolves, let's calculate the overdensity and follow its evolution. The densities evolve as:

$$\begin{aligned} \rho(\theta) &= \frac{M}{(4\pi/3) r^3(\theta)} = \frac{3M}{4\pi A^3} \frac{1}{(1-\cos\theta)^3} \\ \bar{\rho}(\theta) &= \frac{1}{6\pi G} \frac{1}{t^2(\theta)} = \frac{1}{6\pi G B^2} \frac{1}{(\theta - \sin\theta)^2} \end{aligned} \tag{9.14}$$

The ratio of the (9.14) determines the density contrast:

$$1 + \delta = \frac{\rho(\theta)}{\bar{\rho}(\theta)} = \frac{9}{2} \frac{(\theta - \sin\theta)^2}{(1-\cos\theta)^3} \tag{9.15}$$

where we have used the first of (9.7). We can now consider three different stages of the evolution.
Consider the linear regime at early times, where δ is still really small. Expanding

(9.15) to lowest order we get:

$$\begin{aligned}\delta &= \frac{9}{2}\frac{(\theta^3/6-\theta^5/120)^2}{(\theta^2/2-\theta^4/24)^3}-1 \simeq \frac{9}{2}\frac{\theta^6}{36}\frac{8}{\theta^6}\frac{(1-\theta^2/10)}{1-\theta^2/4}-1 \\ &\simeq 1-\frac{\theta^2}{10}+\frac{\theta^2}{4}+\mathcal{O}(\theta^4)=\frac{3}{20}\theta^2=\frac{3}{20}\left(\frac{6}{B}\right)^{2/3}t^{2/3}\equiv\delta_{\rm lin}(t)\end{aligned} \tag{9.16}$$

which is just the linear growth function in a matter dominated universe, a result that we already found in the first chapter.
The turn around point, $\theta=\pi$, is characterized by:

$$\delta(\theta=\pi)=\frac{9\pi^2}{16}-1\approx 4.55 \tag{9.17}$$

It is also useful to ask what the extrapolated linear solution gives at the turn around point, albeit being just an artificial concept:

$$\delta_{\rm lin}(t)=\frac{3}{20}(6\pi)^{2/3}\left(t/t_{\rm max}\right)^{2/3} \implies \delta_{\rm lin}(t_{\rm max})\approx 1.06 \tag{9.18}$$

where we used the second equation of (9.13). This test, whether $\delta_{\rm lin}$ exceeds its limit of validity $\delta_{\rm lin}\ll 1$, is a good test to judge by extrapolation if the perturbation is decoupled from the Hubble flow.
Finally, in the collapse time, the solution diverges:

$$\delta(\theta_{\rm coll}=2\pi)=\infty \tag{9.19}$$

Before discussing the divergence, let's again calculate the result for the extrapolated density contrast in the linear regime, when the matter has collapsed. Using $t_{\rm coll}=2t_{\rm max}=2B\pi$, we find:

$$\delta_{\rm lin}(t_{\rm coll})=\frac{3}{20}(12\pi)^{2/3}\approx 1.69 \tag{9.20}$$

We can therefore say that when the extrapolated linear density contrast reaches 1.69, the region will have collapsed.
The divergence we found in (9.19) is a symptom of having assumed perfect spherical symmetry for the initial perturbation. In fact, our treatment cannot be extended to arbitrarily small values of r. As the mass shell turns around and begins to collapse, particles that are in the mass shell in question can cross the mass shells that were originally inside it and therefore, as was said above, the mass conservation of each shell is not a valid assumption anymore. Indeed, by the time $t=2t_{\rm max}$, all the mass shells have crossed path so many times that the object formed is considered a virialized halo.
In our cosmological model, the halos that form are the dark matter halos, which are the locations where the baryons finally collapse into and form galaxies. In order to estimate the real value of the density of these halos, we can use the virial theorem, which states that, for virialized objects, there is a relation between the average kinetic and potential energy:

$$T=-\frac{1}{2}V \tag{9.21}$$

If we pair this with the conservation of energy, we get that at the turn around point $E = V_{\text{turn}}$, so that after virialization:

$$T_{\text{vir}} + V_{\text{vir}} = \frac{1}{2} V_{\text{vir}} = V_{\text{turn}} \tag{9.22}$$

This means that the radius of the virialized object is halved from the turning point $r_{\text{vir}} = r_{\text{max}}/2$, and the density is $\rho_{\text{vir}} = 8\rho_{\text{turn}}$. We can compare this to the background density at the time of virialization, which we take to be $t_{\text{vir}} \approx t_{\text{col}} = 2t_{\text{max}}$ for simplicity. Since $\bar{\rho} \propto t^{-2}$, this means that $\bar{\rho}_{\text{vir}} = \bar{\rho}_{\text{turn}}/4$.
Finally, the desity of the object is:

$$1 + \delta_{\text{vir}} = \frac{\rho_{\text{vir}}}{\bar{\rho}_{\text{vir}}} = \frac{8\rho_{\text{turn}}}{\bar{\rho}_{\text{turn}}/4} = 32(1 + \delta_{\text{turn}}) = 18\pi^2 \approx 178 \tag{9.23}$$

where we have used (9.17). We found that the density of dark matter halos is therefore around 200 times greater that the average density of the Universe, a result that agrees with numerical simulations.

9.2 The Press-Schechter Formalism

In the standard ΛCDM paradigm of cosmological structure formation, the formation of galaxies begins with the gravitational collapse of the overdense regions into virialized halos of dark matter. Hereafter, the baryons, bound in the potential wells of the said halos, proceed to cool, condense and form galaxies. It is of fundamental importance to understand the formation, the properties and the abundance of these DM halos, if our goal is to ultimately grasp the workings of galaxies.
Press and Schechter (PS) introduced a formalism which made use of the result we found above: whenever a density perturbation in linear theory exceeds the threshold $\delta_c \simeq 1.69$, a virialized halo will have formed.
The first step in this theory is to introduce a linearly evolved smoothed density field by averaging out the contributions below a certain scale R to the density contrast field, so that we obtain a new field, which we shall call $\delta_R(t, \mathbf{x}) \equiv \delta(t, \mathbf{x}; R)$, and is given by a convolution with a specified window function or filter $W(\mathbf{x}, R)$ (which is assumed to be spherically symmetric):

$$\delta_R(t, \mathbf{x}) = \int d^3x' W(|\mathbf{x} - \mathbf{x}'|; R)\delta(t, \mathbf{x}) \tag{9.24}$$

so that in Fourier space it becomes a product: $\delta_R(t, \mathbf{k}) = W(\mathbf{k}; R)\delta(t, \mathbf{k})$.
The purpose of the window function is therefore to weigh in a specific manner (depending on the analytic form) the density field. The most common type of window function is a sphere in real space, the so called top-hat function:

$$W(r; R) = \begin{cases} \frac{3}{4\pi R^3} & r < R \\ 0 & r > R \end{cases} \tag{9.25}$$

with a Fourier transform:

$$W(k; R) = \frac{3}{(kR)^3}(\sin(kR) - kR\cos(kR)) \tag{9.26}$$

The smoothed field $\delta_R(t, \mathbf{x})$ is then, in this case, the average density in spheres of radius R around the point $\mathbf{x}$. The problem with this choice in this window function, however, is that the jump in real space introduces power on all scales in Fourier space, so that it is useful to smoothen the boundary of W, which we do by, for instance, introducing a Gaussian window function:

$$W(x; R) = \frac{1}{(2\pi)^{3/2} R^3} \exp\left(-\frac{x^2}{2R^2}\right) \tag{9.27}$$

with a Gaussian Fourier transform:

$$W(k; R) = \exp\left(-\frac{k^2 R^2}{2}\right) \tag{9.28}$$

Any given window can be labeled by the mean mass contained in it $M(R)$, instead of the size R, where the relation between the two is given by:

$$M(R) \equiv \bar{\rho} V(R) \tag{9.29}$$

where $\bar{\rho}$ is the mean mass density, by convention evaluated at present time, since the mass of the collapsed object is conserved, and $V(R)$ is the volume associated to the convolution window:

$$V(R) = \begin{cases} \frac{4\pi R^3}{3} & \text{Top-hat filter} \\ (2\pi)^{3/2} R^3 & \text{Gaussian filter} \end{cases} \tag{9.30}$$

Since the smoothed density fluctuation field $\delta_R(t, \mathbf{x})$ itself a Gaussian random variable, being a sum of Gaussian random variables, in order to find its probability distribution we need only to know its mean (which is zero) and its variance:

$$\sigma^2(R, t) = \left\langle \delta_R(t, \mathbf{x})^2 \right\rangle = \int d \log k \, \Delta(t, k)^2 |W(k; R)|^2 \tag{9.31}$$

where $\Delta(t, k)^2$ is the dimensionless matter power spectrum. Evaluating this variance for the specific scale $R = 8h^{-1}$ Mpc, gives the parameter σ_8 which is often used as a measure of the amplitude of the (linear) power spectrum. Its value measured today is roughly $\sigma_8 \approx 0.811(6)$. Figure 9.1 shows the variance of the smoothed density field as a function of the mass contained in the averaging volume, for the ΛCDM parameters, and for a Gaussian window function.

As we said above, the Press-Schechter formalism of galaxy formation assumes that the mass inside a region in which the smoothed density fluctuation is greater or equal than a critical value δ_c (which generally depends on the redshift z), corresponds to a virialized object with mass given by $M(R)$, which is given by (9.29) and (9.30). Using this fact we can thus calculate important statistical quantities, like the number of halos in a given mass range. Let $n_h(t, \mathbf{x}, M)$ be the number of halos of mass M at position $\mathbf{x}$ and at time t, and $\bar{n}_h(t, M) \equiv \langle n_h(t, \mathbf{x}, M) \rangle$ be its mean value. Since we said that the smoothed density field is a Gaussian random variable, the

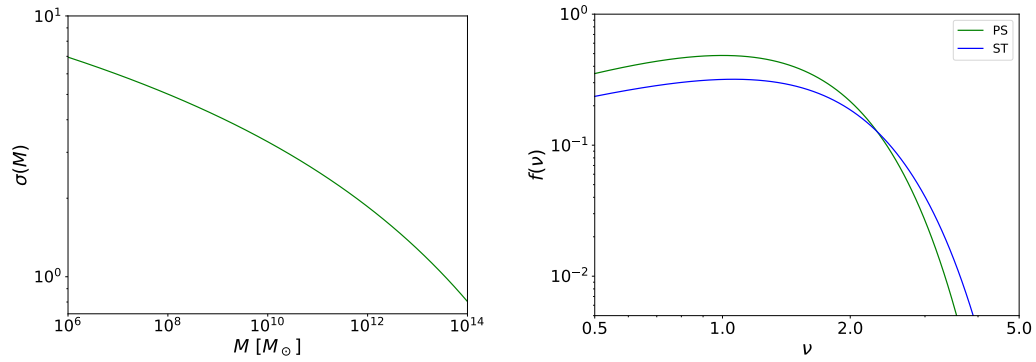

FIGURE 9.1: *Left*: The variance of the (smoothed) overdensity field as a function of mass M contained in the smoothing scale, introduced in (9.31). *Right*: Comparison between the Press-Schechter and the Sheth-Tormen halo multiplicities of eq.s (9.37) and (9.38).

probability that a region of space with mass M (or inside a scale R) has an overdensity δ_M is therefore given by (8.49):

$$\mathbb{P}(\delta_M) = \frac{1}{\sqrt{2\pi\sigma(M)^2}} \exp\left(-\frac{1}{2}\frac{\delta_M^2}{\sigma(M)^2}\right) \tag{9.32}$$

where $\sigma(M)^2$ was just introduced in (9.31), and the time dependence is suppressed. Now the probability for a region with overdensity δ_M to exceed the critical value δ_c is:

$$\mathbb{P}(\delta_M > \delta_c) = \int_{\delta_c}^{\infty} d\delta_M\, \mathbb{P}(\delta_M) = \int_{\nu}^{\infty} dx\ \exp\left(-\frac{x^2}{2}\right) = \frac{1}{2}\mathrm{erfc}\left(\frac{\nu}{2}\right) \tag{9.33}$$

where $\mathrm{erfc}(x)$ is the complementary error function, and$\nu(M) \equiv \delta_c/\sigma(M)$ is the "peak height", the height of the threshold in units of standard deviations of the smoothed density distribution. Note that in this model, the collapse of a region of mass M is defined so that it occurs when the smoothed density $\delta_R(t, \mathbf{x})$ equals δ_c on the appropriate (mass) scale.

Note that in the standard ΛCDM model, the variance $\sigma(M)$ is a monotonically decreasing function of the scale M, which implies that small scale fluctuations are the first to collapse. This type of structure formation is called "bottom up", because large-scale structure are formed when smaller and smaller objects coalesce into larger ones.

The Press-Schechter approach has a problem however, which is clear when looking at the result (9.33). In fact, in the limit $R \to 0$, $\mathbb{P}$ should give the fraction of all mass in virialized objects, however, since $\sigma(M)$ becomes arbitrarily large when $R \to 0$ and $\mathrm{erfc}(0) = 1$, in this limit we get that only half of the mass collapses into halos, which is not consistent with numerical simulations. This issue arises because this approach does not account for the fact that while at a particular smoothing scale R the smoothed density field $\delta_R(t, \mathbf{x})$ could be smaller than δ_c, it may well be possible that it is bigger than the critical value at a bigger smoothing scale $R' > R$; then it would seem natural that this larger volume should collapse to form a virialized object, and in the meanwhile swallow up the smaller parts within it. This effect clearly has the consequence of increasing our result in (9.33).

Press and Schechter "solved" this problem by multiplying (9.33) by a factor of 2, $\tilde{\mathbb{P}} = 2\mathbb{P}$, even though the reason why it should have been exactly a factor of 2 was far from convicing. In the literature, this issue of the existence of regions below a certain threshold on a particular scale, which are above threshold on a bigger scale, is called the "cloud-in-cloud" problem.
Some time later, Bond, Cole and Efstathiou and Kaiser, introduced an extension of the PS theory, using excursion set theory, that explained the factor of 2.
Muliplying by the factor of 2, one finds that the probability that a halo formed in the mass range $[M, M + dM]$ is:

$$P([M, M + dM]) = \left|\tilde{\mathbb{P}}(\delta_{M+dM} > \delta_c) - \tilde{\mathbb{P}}(\delta_M > \delta_c)\right| \approx -\frac{d\tilde{\mathbb{P}}}{dM} \tag{9.34}$$

The abundance of halos of mass M, called the "halo mass function", is then obtained by multiplying (9.34) by the maximum number of such halos in a region of mean density $\bar{\rho}$, which is given by $\bar{\rho}/M$:

$$\frac{d\bar{n}_h}{dM}(z, M) = -\frac{\bar{\rho}}{M}\frac{d\tilde{\mathbb{P}}}{dM} = -\sqrt{\frac{2}{\pi}}\nu(M)\exp\left(-\frac{\nu(M)^2}{2}\right)\frac{\bar{\rho}}{M^2}\frac{d\log\sigma(M)}{d\log M} \tag{9.35}$$

Note that this equation contains a time (redshift) dependence through $\sigma(M)$, where it is found in the power spectrum, as is written in (9.31). Using the ΛCDM linear matter power spectrum, we find that the mass function is a power law for small masses (high ν), while it has an exponential fall-off at large mass scales (low ν). Note that through this formalism, we can understand the way nonlinear structure develops: looking at (9.35), we see that halos with mass M can only form in significant number only when the argument in the exponential is not too large, so that $\sigma(M) \gtrsim \delta_c$; in fact, by defining a characteristic mass $M^\star$ by:

$$\sigma(M^\star, t) = \delta_c \tag{9.36}$$

only halos with mass $M \lesssim M^\star$ can have formed in large number before time t.
The function of the peak height ν appearing in the mass function is called the "halo multiplicity":

$$f_{PS}(\nu) = \sqrt{\frac{2}{\pi}}\nu\exp\left(-\frac{\nu^2}{2}\right) \tag{9.37}$$

The PS mass function captures all the essential features of structure formation, but it disagrees with the results of numerical $N-$body simulations, in particular it underpredicts the abundance of rare high-mass halos by about a factor of 10, and overpredicts that of low-mass halos by a factor of 2. One way to obtain a better fit with the simulations would be to realize that the PS formalism assumes that the collapse of density perturbations is described by the spherical collapse model. However, it can be shown that the collapse of overdensities in a Gaussian density field is in general ellipsoidal, rather than spherical.
Sheth and Tormen, considering these effects, proposed a new fitting function for the halo multiplicity:

$$f_{ST}(\nu) = A\sqrt{\frac{2}{\pi}}[1 + (a\nu^2)^{-p}]\sqrt{a\nu^2}\exp\left(-\frac{a\nu^2}{2}\right) \tag{9.38}$$

where $A = 0.32$, $a = 0.75$ and $p = 0.3$. Figure 9.1 shows a comparison between the halo multiplicities of Press-Schechter and Sheth-Tormen.

9.3 Excursion Set Theory

As was discussed in the previous section, the Press-Schechter derivation of the halo mass function (9.35) is plagued by a problem: it does not allow for the possibility that $\delta_R(\mathbf{x})$ can be smaller than the critical density δ_c at that smoothing scale R, but could be larger at a bigger scale $R' > R$. In this case it is only natural to assume that the larger volume R' would swallow up the smaller scale and form a virialized object. This problem was therefore solved by increasing, by a factor of 2, the probability in (9.33).
To really solve the "cloud-in-cloud" problem, using what's called the excursion set formalism, we would need to compute the largest value of the smoothing scale R for which the $\delta_R(\mathbf{x})$ exceeds the threshold δ_c.
We first consider the smoothed density field $\delta_R(\mathbf{x})$ now not as a function of the comoving coordinate $\mathbf{x}$, but as a function of the scale R, so that we can just write $\delta(R)$. For very large scales, we know that $\delta_R(\mathbf{x}) \ll \delta_c$, so the probability that the region lies above the boundary is vanishingly small. As we decrease R, the variance becomes larger and therefore the threshold will be crossed at some point, which is the first up-crossing of the barrier. The problem is therefore to compute the equation of motion of $\delta(S)$, where we denoted $\sigma^2(R) \equiv S$, and used the fact that the variance is a bijective function of the scale, so that we can interchange them freely. Note that each location $\mathbf{x}$ now corresponds to a trajectory $\delta(S)$, which reflects the value of the smoothed density field at that point in space, when smoothed over a scale S.
Throughout the discussion it is important to remember that the variance is a (monotonically) decreasing function of the scale R, and as a consequence we have that bigger S means smaller M and viceversa. Therefore the problem is to calculate the first up-crossing between the values S and $S + dS$.
Consider a certain value $S = S_1$ as the starting point, with $\delta(S_1) \equiv \delta_1 < \delta_c$. For a given change in S space, we may have a certain probability of reaching the value δ_2 after an increment $\Delta S = S_2 - S_1 > 0$ (decreasing the scale R or M), which in general depends both on the size of the step ΔS and the value of $\delta(S)$ on other scales. If the probability distribution that we are looking for depends on the value of the density field on other scales, the equations governing its motion are non-trivial, therefore we solve this problem by assuming a sharp-k smoothing window function, used to define (9.31):

$$W(k, R) = \Theta(1 - kR) \tag{9.39}$$

The problem with this sharp transition in Fourier space is that the integration of $W(\mathbf{x}, R)$ over all space diverges, so it is not straightforward to associate to it a well defined volume; formally, we assign a volume as if we were working with a tophat filter in real space:

$$V(R) = \frac{4}{3}\pi R^3 \tag{9.40}$$

In any case, using this filter, we can see that decreasing the filter scale just corresponds to adding a set of independent Fourier modes to the smoothed density. These modes, of course, have not played a role in determining $\delta(S)$ at other smoothing scales, thus the probability of a jump in density $\Delta\delta$ associated with the jump in

variance ΔS is a Gaussian with zero mean and variance equal to ΔS, independently of the starting point δ_1 in S space.
What one finds is the following: the trajectory $\delta(S)$ as we change the smoothing scale executes a Brownian random walk, of which examples are shown below:

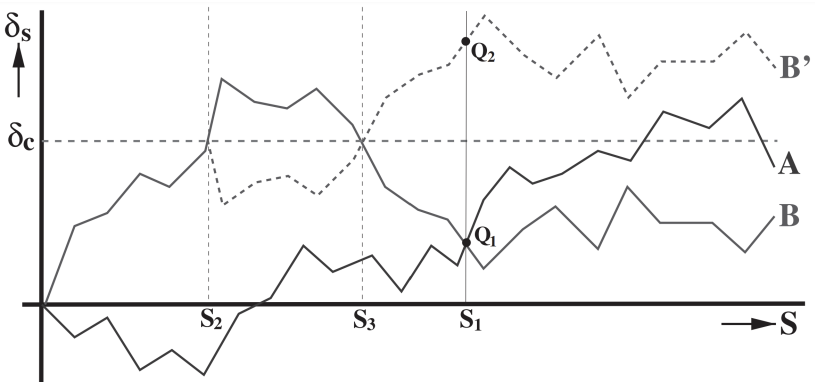

FIGURE 9.2: Gaussian random walks of the smoothed density field δ_S as the mass scale S is varied, governed by the diffusion equation (9.45). Three relevant trajectories A, B and B′ are show, where B′ is the mirror of B from the point (δ_c, S_2) onwards. Since the transition probability is a Gaussian, B and B′ are equally probable.

In light of this, the goal of excursion set then becomes to calculate the largest smoothing scale R, or equivalently M, at which the random walk pierces the barrier δ_C. This line of reasoning introduces naturally the missing factor of 2 from the Press-Schechter theory in the following way. Consider the two random walks A and B in Figure 9.2 and a mass scale M_1 with an associated variance S_1, where the continuous vertical line is positioned. The Press-Schechter ansatz assumes that the fraction of trajectories that have $\delta_1 > \delta_c$ is the fraction of mass elements in collapsed objects with a mass $M > M_1$. The problem is that, for instance, the trajectory B is above the critical density in between the scales S_2 and S_3, so that it should be found as part of a halo with mass $M > M_3 > M_1$. The Press-Schechter theory does not account for these type of trajectories, like A and B, and therefore some mass is missing, i.e. the factor of 2. A correction for this can be made with a heuristic approach, if one realizes that the trajectory from (δ_2, S_2) to (δ_1, S_1) is equally as likely as the trajectory B′, the mirror of B for $S \geq S_2$, since the transition probability is a Gaussian which depends only on the size of the step, as we remarked above. This implies that each trajectory that is missed by the PS theory, corresponds to a mirrored trajectory whose smoothed density field is above the threshold at the desired mass scale, and therefore the fraction of mass in halos with mass greater that the said scale is given by twice the fraction of actual trajectories above threshold. This reasoning gives a natural explanation for the missing factor of 2.
More mathematically, we know that in the case of a sharp-k filter, the transition probability from δ_1 to $\delta_2 = \delta 1 + \Delta\delta$ is:

$$\Pi(\delta_2, S_2)d\delta_2 = \Psi(\Delta\delta; \Delta S)d(\Delta\delta) \tag{9.41}$$

where the Ψ only depends on the size of the jump ΔS, as remarked above:

$$\Psi(\Delta\delta; \Delta S)d(\Delta\delta) = \frac{1}{\sqrt{2\pi}\Delta S} \exp\left(-\frac{(\Delta\delta)^2}{2\Delta S}\right) d(\Delta\delta) \tag{9.42}$$

Since we now want to find the evolution equation of $\Pi(\delta, S)$, we start with the relationship between it and its value at a subsequent step $\Pi(\delta, S + \Delta S)$:

$$\Pi(\delta, S + \Delta S) = \int d(\Delta\delta)\Psi(\Delta\delta; \Delta S)\Pi(\delta - \Delta\delta, S) \tag{9.43}$$

Now we Taylor expand (9.43), keeping terms only up to $(\Delta\delta)^2$ and integrating to find:

$$\frac{\partial\Pi}{\partial S} = \lim_{\Delta S \to 0}\left(\frac{\langle(\Delta\delta)^2\rangle}{2\Delta S}\frac{\partial^2\Pi}{\partial\delta^2} - \frac{\langle\Delta\delta\rangle}{\Delta S}\frac{\partial\Pi}{\partial\delta}\right) \tag{9.44}$$

Using the fact that the transition probability is Gaussian with zero mean $\langle\Delta\delta\rangle = 0$ and variance $\langle(\Delta\delta)^2\rangle = \Delta S$, we simplify things further:

$$\frac{\partial\Pi}{\partial S} = \frac{1}{2}\frac{\partial^2\Pi}{\partial\delta^2} \tag{9.45}$$

which is precisely the diffusion equation. In order to calculate the first up-crossing of $\delta(S)$ at δ_c, we wish to solve (9.45) for trajectories that have a starting point $(\delta(S_0), S_0)$ and reach a point $(\delta(S), S)$ without exceeding a critical value δ_c at smaller S. This is equivalent to the diffusion equation with an absorbing barrier at $\delta(S) = \delta_c$. This problem has the following solution:

$$\Pi(\delta, S) = \frac{1}{\sqrt{2\pi}\Delta S}\left(\exp\left(-\frac{(\Delta\delta)^2}{2\Delta S}\right) - \exp\left(-\frac{(2(\delta_c - \delta_0) - \Delta\delta)^2}{2\Delta S}\right)\right) \tag{9.46}$$

where $\Delta S = S - S_0$ and $\Delta\delta = \delta - \delta_0$. We see that the first term in (9.46) is the Gaussian probability distribution that represents the points above threshold at S, while the second term accounts for the trajectories that have been removed because they crossed above threshold at a bigger scale $S' < S$, but would have crossed back below by S, which are the trajectories that the Press-Schechter formalism fails to take into account.

At this point, from the discussion it is clear that the fraction of trajectories that have crossed the threshold at or prior to a certain scale $S(M)$ is the complement of the sum of (9.46) until δ_c:

$$\mathbb{P}(S) = 1 - \int_{-\infty}^{\delta_c}\Pi(\delta, S)d\delta = \operatorname{erfc}\left(\frac{\delta_c - \delta_0}{\sqrt{2}\Delta S}\right) \tag{9.47}$$

Taking now the starting point to be $(\delta_0, S_0) = (0, 0)$ at some large value of the smoothing scale, the above equation yields precisely the Press-Schechter mass function in the following way. The differential probability for a first piercing of the threshold is:

$$\begin{aligned} f(S)dS \equiv \frac{d\mathbb{P}(S)}{dS}dS &= \left(\int_{-\infty}^{\delta_c}\frac{\partial\Pi}{\partial S}d\delta\right)dS = \frac{1}{2}\frac{\partial\Pi}{\partial\delta}\Bigg|_{-\infty}^{\delta_c}dS \\ &= \frac{\delta_c}{\sqrt{2\pi}S^{3/2}}\exp\left(-\frac{\delta_c^2}{2S}\right)dS \end{aligned} \tag{9.48}$$

Now the fraction of mass in collapsed objects in a narrow interval of masses is found by changing the variable to M:

$$\frac{d\mathbb{P}(M)}{dM} = \frac{1}{\sqrt{2\pi S}} \frac{\delta_c}{S} \left| \frac{dS}{dM} \right| \exp\left(-\frac{\delta_c^2}{2S} \right) \tag{9.49}$$

The mass function is given by multiplying (9.49) by $\bar{\rho}/M$, yielding the Press-Schechter result, without the ad-hoc factor of two.
Note that in the excursion set formalism, the sharp-k filter Fourier window function only serves to simplify the calculations, but in principle other filters can be used at the cost of increasing mathematical complexity. As we noted above, this is because the steps in S space are not independent anymore, so it is necessary to compute the trajectory all at once to account for the correlations. It is important however to underline that the lack of correlations is not a prediction of excursion set theory, but only a simplifying assumption.

9.4 Halo Formation In the Excursion Set Theory

Excursion set theory is a powerful tool to calculate the properties of the progenitors which give rise to a given class of collapsed objects. For example, one can calculate the mass function at $z = 5$ of these halos which by $z = 0$ end up in a massive cluster-sized halo of mass $10^{15} M_\odot$. Basically, excursion sets allow us to track how a dark matte halo assembled its mass via mergers of smaller mass halos.
Since we need to make explicit the time dependence, we will write the evolution of the smooothed density field as $\delta(S, a) = D(a)\delta(S, a = 1)$ where $D(a)$ is the linear growth factor (8.93) (normalized so that $D(a = 1) = 1$).
As we've seen in the last section, the formulas depend only on the ratio $\delta_c/\sigma(M)$, so that it is often easier to extract the growth factor from the rms variance and envision the height of the critical density threshold as a function of time. In that way, collapse at a scale factor $a' \neq 1$ corresponds to the $a = 1$ density fluctuation penetrating a barrier of height $\delta_c(a') = \delta_c/D(a')$.
While at first the possibility of using excursion sets to understand the formation histories of halos was discussed at first, only later was the problem studied more formally; we will write:

$$\omega(a) \equiv \delta_c/D(a) \tag{9.50}$$

Consider now a spherical region of mass M_1 corresponding to a variance $S_1 = \sigma^2(M_1)$, and indicate $\omega_1 = \omega(t_1) = \delta_c/D(t_1)$ so that this object collapses at time t_1. We are now interested in the fraction of M_1 that was already in collapsed objects of a certain mass at an earlier time $t_2 < t_1$. To calculate this, we calculate the probability of a random walk originating at (δ_1, S_1) to execute a first up-crossing of the barrier $\omega(t_2)$ at $S = S_2 > S_1$, corresponding to a mass scale $M_2 < M_1$. This is exactly the same problem as above, although with a shift in the origin, which we now take to be (δ_1, S_1), hence the probability we are looking for is the following:

$$f(S_2, \omega_2 | S_1, \omega_1) dS_2 = \frac{1}{\sqrt{2\pi}} \frac{\Delta\omega}{\Delta S^{3/2}} \exp\left(-\frac{(\Delta\omega)^2}{2\Delta S} \right) dS_2 \tag{9.51}$$

where $\Delta\omega = \omega_2 - \omega_1$ is the difference in barrier heights, $\Delta S = S_2 - S_1$. According to the excursion set theory, (9.51) is the conditional probability that a trajectory pierces the barrier ω_2 in an interval of width dS_2 about S_2 on the condition that the trajectory first pierces ω_1 at S_1, or in other words, it gives the fraction of mass elements in $(\delta_1, M1)$ that were in collapsed objects of mass M_2 at the earlier time t_1.
Note that in this section, the interpretation of (9.51) is different from the one we gave above, where the initial condition was fixed at a large scale $S_1 \ll S_2$. Here, the two barriers represent the critical density at two different times, so that the change in barrier height $\Delta\omega$ represents a shift in time.

The Conditional Mass Function

Using (9.51), a quantity of interest that descends from it is the "conditional mass function". Given a halo of mass M_1, one can obtain the average number of progenitors at t_2 in the mass range $(M_2, M_2 + dM_2)$ that have merged with it by the time t_1. This is done by converting (9.51) from mass weighting to number weighting:

$$\frac{dn(M_2, \omega_2|M_1, \omega_1)}{dM_2} dM_2 = \frac{M_2}{M_1} f(S_2, \omega_2|S_1, \omega_1) \left|\frac{dS_2}{dM_2}\right| dM_2 \tag{9.52}$$

The following figures show the conditional mass function for standard ΛCDM power spectra, for four halo masses at three different redshifts $z_2 = 1$, $z_2 = 1/4$, $z_2 = 3$:

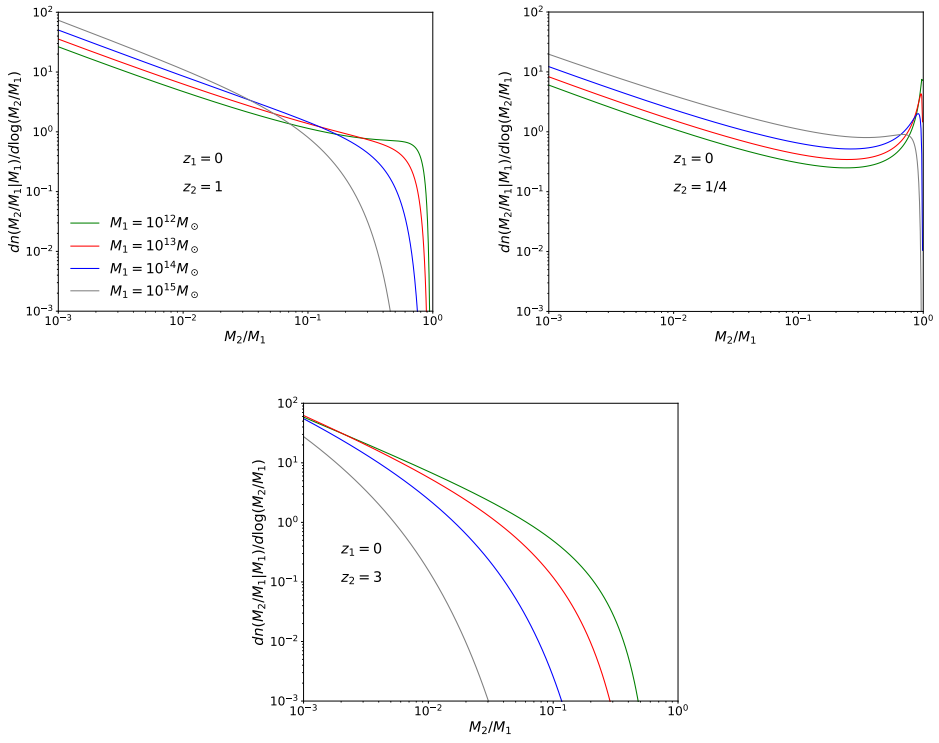

FIGURE 9.3: The conditional mass function of equation (9.52) at three different redshifts, for four different halo masses

The evolution towards more fragmentation as the redshift z_2 of the smaller objects increases is evident in Figure 9.3. Also notice that more massive halos tend to

fragment into smaller sub-units more quickly. The progenitor mass functions (9.52) have been tested by numerical simulations, which show that although the agreement is good when the time difference $\Delta\omega$ between the epochs is small, the conditional mass function significantly underestimates the mass fraction in high mass progenitors for relatively large $\Delta\omega$.

Halo Accretion Rates

Another possibility is that of following the growth of a dark matter halo in time through mergers/accretion from lower mass objects. In order to do this, we ask what is the probability that some object with a certain mass will be incorporated into a bigger halo at some later time.
Therefore, we clearly need to ask the opposite question compared to the previous section, so that we start with an already existing object of mass M_2 at scale S_2 (and formed at time ω_2), and look for the posterior probability:

$$\begin{aligned} f(S_1,\omega_1|S_2,\omega_2)dS_1 &= \frac{f(S_2,\omega_2|S_1,\omega_1)f(S_1,\omega_1)}{f(S_2,\omega_2)}dS_1 \\ &= \frac{1}{\sqrt{2\pi}}\frac{\omega_1(\omega_2-\omega_1)}{\omega_2}\left(\frac{S_2}{(S_1(S_2-S_1))}\right)^{3/2}\exp\left(-\frac{(\omega_1 S_2-\omega_2 S_1)}{2S_1S_2(S_2-S_1)}\right)dS_1 \end{aligned} \tag{9.53}$$

where $f(S,\omega)$ is given by (9.48). According to the formalism we developed, this formula is the conditional probability that a halo of mass M_2 at time t_2 is incorporated into a halo with a mass between $M_1 > M_2$ and $M_1 + dM_1$ at a later time $t_2 > t_1$. If we then set $M_1 \equiv M_2 + \Delta M$ and $t_1 = t_2 + \Delta t$, equation (9.53) gives the probability for the halo to gain a mass ΔM by merging/accretion in the time interval Δt.
Finally, the rate at which a halo with initial mass M_2 transits to a halo with mass between M_2 and $M_1 = M_2 + \Delta M$ per unit ΔM and per unit Hubble time is:

$$\frac{d^2R}{d\log\Delta M d\log a} = \sqrt{\frac{2}{\pi}}\frac{\Delta M}{M_1}\frac{\omega/\sigma(M_1)}{(1-S_1/S_2)^{3/2}}\exp\left(-\frac{\omega^2(S_2-S_1)}{2S_1S_2}\right)\left|\frac{d\log\omega}{d\log a}\right|\left|\frac{d\log\sigma}{d\log M_1}\right| \tag{9.54}$$

where we took the limit $\omega_2 - \omega_1 \to 0$ to give a quantity per unit time, i.e. the rate. Note that a change in mass ΔM can be, in principle, due to a cumulative effect of multiple mergers. However, the accretion rate is only interested in infinitesimal shifts in time, so that the change in mass is most likely due to a single event.
Thus the formula (9.54), the "halo accretion rate", gives the number of mergers per unit of Hubble time of a halo with mass M_2 with a smaller object with mass M_1. A quantity of interest is also (9.54), but weighted by the shift in mass we are considering, the "mass-weighted halo accretion rate", which is basically the rate of mass increase.
The accretion rates and the mass accretion rates for the standard ΛCDM cosmology are shown in the following figures:

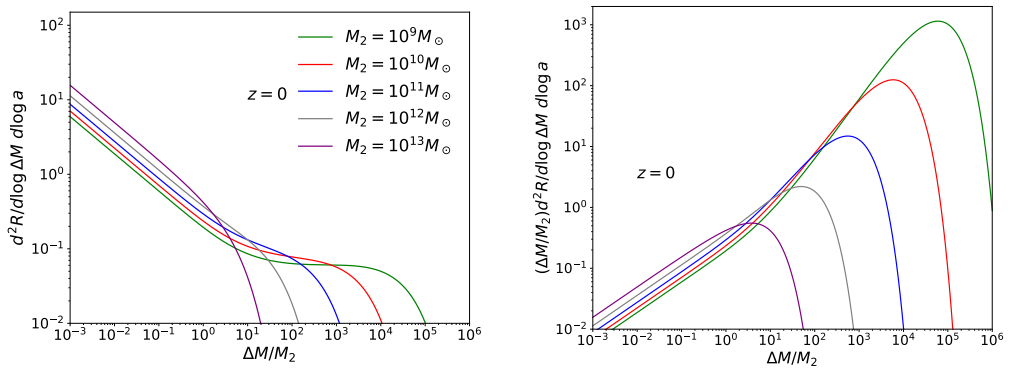

FIGURE 9.4: Halo accretion rates per logarithmic interval of mass change and Hubble time for five different initial masses at $z = 0$. *Left*: the probability for a change in ΔM as a function of the mass shift itself as given in (9.54). *Right*: the fractional mass accretion rate given by (9.54) times $\Delta M/M_2$.

Note the following points. The halo mass accretion rate on the left of Figure 9.4 diverges for small increases, which means that the smaller mergers are most frequent in general. However, the average mass increase rate converges at $\Delta M \to 0$ and is dominated by the more infrequent high mass mergers, i.e. the halos get most of their mass increase by these larger mergers compared to accreting smaller objects, albeit in very large quantities.
Finally, the "total mass accretion rate" is defined by integrating (9.54) over $\log \Delta M$, which gives the total mass increase per unit of Hubble time of an object with initial mass M_2. The following figure computes this integral and shows it in unit of halo mass and at three different redshifts $z = 0$, $z = 0.5$ and $z = 1$:

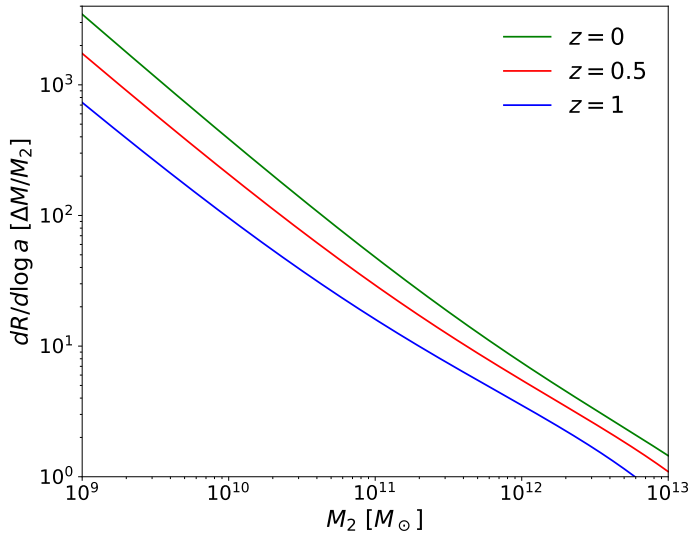

FIGURE 9.5: Total mass increase rate, defined as the integral of (9.54) over $\log \Delta M$, in units of halo mass as a function of initial masses, for three different redshifts $z = 0$, $z = 0.5$, $z = 1$

Figure 9.5 shows clearly that smaller halos increase their mass many times over

the span of a Hubble radius, but this increase gets attenuated as we get farther from the present epoch (increasing redshift).
An inclusive interpretation of Figures 9.4 and Figure 9.5 is that halos residing in the smaller mass scales tend to increase their mass more rapidly, mostly through mergers with larger objects, whereas bigger halos accrete more slowly and predominantly objects that are smaller in size.