Lecture Notes

Cosmology

Rishi Khatri
Tata Institute of Fundamental Research
khatri@theory.tifr.res.in

Last Modified October 16, 2024

Contents

Chapter 0

How the Universe was discovered

> "Space," it says," is big. Really big. You just won't believe how vastly, hugely, mindbogglingly big it is. I mean, you may think it's a long way down the road to the chemist's, but that's just peanuts to space, listen..."
>
> ---
>
> The hitchhiker's guide to the galaxy
> - Douglas Adams

The year 2020 marks the 100 years since the *Great Debate*[1] between Harlow Shapley and Heber Curtis. The subject was nothing less than the nature of the Universe. More precisely, the question being debated was whether there were galaxies other than our own, the Milky Way, in the Universe or the Milky Way was the entire Universe.

Thomas Wright [1] was the first astronomer to give the correct explanation for the appearance of the Milky Way in the night sky [2]. He argued that the diffuse *milky* appearance of the Milky Way was because of the large number of stars in that region of the sky, too numerous to be seen individually. This meant that the distribution of stars in the Universe was not isotropic but must be concentrated in a disc and we must lie in the plane of the disk, since we see the disk only edge on as a thin band on the sky. There were however other luminous clouds in the sky, the so called nebulae. Wright and Immanuel Kant [3] in the 1750s speculated that the other faint nebulae in the sky, such as the Andromeda, could also be systems of stars or *island universes* outside the Milky Way [4]. However, at the time there

[1] `https://apod.nasa.gov/diamond_jubilee/debate20.html`

was no way to prove this hypothesis.

By the time of the great debate, there were measurements of the light spectrum of the spiral nebulae which were remarkably similar to the integrated spectrum of the stars in the Milky Way. There were also observations of novae, thermonuclear explosions of stars, in the spiral nebulae [5]. However, all this was indirect evidence and left room for doubt in the absence a reliable measurement of distances to the nebulae which could put them firmly inside or outside the Milky Way. The measurements of the spectra of the spiral nebulae by Slipher [6] showed that most of the spiral nebulae had enormous speeds, much faster than any star in the Milky Way, and were receding away from us. These observations were hard to interpret at the time and only served to add to the confusion.

In 1912, Henrietta Leavitt found that the bright variable stars in the Magellanic clouds, the Cepheid variables, showed a remarkable correlation between their period of oscillation and brightness [7]. Thus, by just measuring the period of oscillation of these stars, we could know their absolute luminosity and measure the distance to them. By the time of the great debate, these stars were measured only in the Milky Way and the small and the large Magellanic clouds, the two satellite galaxies of the Milky Way. In 1924 Hubble resolved and measured the period and brightness of the Cepheid variable stars in the nearby spiral nebulae [8, 9]. Hubble's measurement of their distance put them further than any estimate of the size of the Milky Way and thus establishing them firmly to be separate systems of stars or galaxies, just like the Milky Way. The Cepheid variable stars even today form an important part of the cosmic distance ladder used to measure distances to the furthest galaxies in the Universe.

Exercise 1

List three (or more) things that Shapley got right? What three (or more) things did Curtis got right? What did they get wrong? You can read the original papers as well as a shorter transcript of the debate at the NASA website `https://apod.nasa.gov/diamond_jubilee/debate20.html`. There is also an excellent modern analysis of the debate by Virginia Trimble [10] available from NASA ADS website.

0.1 The cosmic distance ladder

How to accurately accurately measure the distances to cosmological objects, in particular distant galaxies, is even today one of the most important areas of research. There are many ways to measure cosmic distances, most of them however require calibration leading to the concept of the distance ladder.

Parallax distance - The distance to the nearby stars can be measured directly using parallax. This is the most straghtforward way of measuring the distances. Measure the position of the star when the angle made by the line joining earth to star with the line joining the Sun to the star is maximum. Then measure the position of the star again after six months. Earth would have moved a distance of 2 astronomical units (au), the average distance between Earth and Sun, to the other side of the Sun in six months. The change in the position of the star is $2\theta_p$, where θ_p is defined to be the *parallax*. The unit of distance parsec (pc) used in astronomy is abbreviation for 1 *par*allax *sec*ond and is defined as the distance at which the parallax of a star is 1 arcsec and 1 pc = 3.262 light years. Therefore the parallax, $\theta_p = \mathrm{pc}/d$, decreases with increasing distances and our ability to measure distances in this way is limited by how accurately we can measure the angular positions. The best instrument today to measure the angular positions of a large number of stars is the Gaia satellite, which can achieve an angular resolution of $\sim$ 100 μarcsec. Therefore we can measure the distances with the parallax method only to about $1\,\mathrm{au}/\theta_p \sim 10\,\mathrm{kpc}$, well within our Galaxy. Note that we are measuring the distances with respect to the Earth-Sun distance. Therefore, the calibration that goes into the parallax distances is a prior measurement of the Earth-Sun distance or the definition of 1 au.

Luminosity distance - If we know the luminosity (L), the total energy emitted per unit time in an electromagnetic band, of a star and measure its flux (F), the total energy per unit time per unit area received at a detector on Earth, we can find the distance (d) assuming conservation of photons i.e. no photons are absorbed (or extra photons emitted) between the star and us, $F = L/(4\pi d^2)$. However, what we need to use this method is prior knowledge of the luminosity of the star. In general, the stars come in different masses, at different stages in there evolution and the luminosities encompass a wide range. We have two requirements in order to use the luminosity distance method:

1. We want to be able to calibrate the luminosities of the stars against other observables which do not depend on the distance.

2. The stars should have high luminosity so that we can observe them at great distances.

One example of an observable that does not depend on the distance is the spectrum of a star. If can find a correlation between the spectrum of a star and its luminosities, we could tell its luminosity by measuring its spectrum. Such correlations do exist and this was one of the methods of determining distances that Shapley and Curtis relied on in their debate.

Another example is pulsating stars such as Cepheid variables. These are massive stars with luminosities upto $10^5 L_\odot$, where $L_\odot$ is the solar luminosity, which are going through a period of instability and exhibit radial oscillations. These oscillations are observed as periodic dimming and brightening of the stars. However, for any star we can only measure the flux at Earth directly and not luminosity. To get the luminosity for a star from the measured flux we should already know its distance. We can measure parallax distance to the Cepheid stars within our Galaxy as well as their period. The period of oscillations of these stars is found to be tightly correlated with the luminosity with more luminous stars having longer periods. This means that once we have calibrated the period-luminosity relation i.e. we have luminosity as a function of period for a sample of stars for which we can measure both period and distance or luminosity, for any new star that is also a Cepheid variable we can get the luminosity from the period-luminosity relation. Thus combination of parallax distances to nearby Cepheid stars together with the period luminosity relation forms a two-rung distance ladder that we can use to measure distances to galaxies too far for the parallax method alone. The Cepheid variable thus are *standard candels*.

Recent measurements from Gaia satellite give the following average relation [11]

$$\langle M_V \rangle = (-2.67 \pm 0.17) \log P - (1.58 \pm 0.16) , \tag{1}$$

where ,P is the period in units of days, M_V is the visible band absolute magnitude and angular brackets indicate average. The magnitude system is a logarithmic scale used in astronomy for observed flux with the magnitude m defined as [12]

$$m = -2.5 \log \left(\frac{F}{F_0} \right), \tag{2}$$

where $F_0 = 2.518 \times 10^{-5}$ ergs/s/cm^2 is the reference flux and F is the observed flux. The absolute magnitude M is the magnitude of the star when it is placed at a distance of 10 pc from us.

$$M = -2.5 \log \left(\frac{L}{4\pi (10 \text{ pc})^2 F_0} \right) \tag{3}$$

$$= -2.5 \log \left(\frac{L}{L_0} \right) \tag{4}$$

where L is the luminosity of the star and $L_0 = 3.0128 \times 10^{35}$ ergs/s is reference luminosity.

Exercise 2

Oscillation Period of Stars: Show by dimensional analysis that the time period of oscillations of stars $P \propto \frac{1}{\sqrt{\rho}}$, where ρ is the mass density. Assume a toy model of star with constant density.

Exercise 3

Distance ladder: We observe a Cepheid variable star near us with an oscillation period P and with total energy flux in visible light at the maximum luminosity to be 3.2×10^{-5} ergs/s/cm^2. The maximum variation in the angular position of the star, when measured 6 months apart (also known as twice the parallax), is found to be 0.02 arcsec. A Cepheid is observed also in the Andromeda galaxy with a period of 1.5P. The flux measured from the Andromeda Cepheid star is 8.2×10^{-13} erg/s/cm^2. What is the distance to Andromeda ?
Hint: Distance of Earth to Sun is 1 au = 1.496×10^{13} cm.

The Cepheid variable stars, although quite bright compared to the Sun, are still not bright enough for cosmological distances. We can go upto few tens of Mpc with the Cepheid variables. To go further, we need stars which are brighter still or rather exploding stars. Most of the massive stars, with mass larger than $\sim 8M_{\odot}$, where $M_{\odot}$ is the mass of the Sun, end their lives in spectacular explosions called the core collapse supernovae. However most supernovae lack the first property, an observable that is tightly correlated with the luminosity. It turns out that there is particular kind of supernova, called the Type 1a supernova, that has a tight correlation between the shape of the light curve, i.e. how the brightness falls as a function of time after the explosion, and the luminosity of the supernova. The mechanism for a Type 1a supernova is very different from a core collapse supernova. Type 1a supernova happens when a white dwarf star in a binary system accreting matter from the companion star, reaches beyond the Chandrasekhar limit of $1.4M_{\odot}$. At this point the material of the white dwarf, mostly carbon, nitrogen and oxygen, undergoes nuclear fusion. The thermonuclear process runs away igniting more and more material until the whole star explodes. The companion star can be a normal star, in which case there is gradual accretion of material onto the white dwarf until the Chandrasekhar limit is breached. A second possibility is that the second star is also a white dwarf and the supernova happens when the two stars merge together.

Whatever be the dominant mechanism, it has emprically been shown that there is a tight correlation between the light curves of the Type 1a supernovae and their peak luminosities [13]. Thus the third rung of the cosmic distance ladder is composed of the supernovae calibrated in the nearby galaxies to which we have accurate distances using the Cepheid variable stars. Once we have the calibration relation, we can observe and measure distances half-way accross the Universe, upto many Gpc, with the Type 1a supernovae.

0.2 The Hubble law

Slipher's measurement of velocities of nearby galaxies using Doppler shift of spectral lines had already indicated that most of the galaxies were receding away from us. With the Hubble's measurement of distances to the galaxies combined with the already existing data about their velocities, a pattern began to emerge. It seemed not only most of the galaxies were moving away from us but further away a galaxy was, faster it was moving [14],

$$v = H_0 R, \tag{5}$$

where v is the radial velocity of the galaxy, R is the distance to the galaxy and H_0 is the Hubble constant, usually measured in units of km/s/Mpc. This is known as the Hubble law. It turned out that this was exactly the newly proposed theory of general relativity by Einstein predicted. Friedmann had already found cosmological solutions for homogeneous isotropic universes to the Einstein's equations which indicated that the Universe should either expand or contract with time [15, 16]. Lemaitre had already tried to connect the Friedmann's solutions to Slipher's observations indicating an expanding Universe [17]. However, it were the more reliable measurements of distances using the Cepheid variable stars by Hubble which finally established the Hubble law and convinced Einstein to abandon his attempts for a static Universe and accept an expanding Universe. An interesting historical account of the discovery of the expansion of the Universe is given by Trimble [18].

Hubble constant has historically turned out to be a tricky quantity to measure precisely and even today one of the most pressing problems in cosmology is a precise determination of H_0. In particular, direct measurement of H_0 in the local Universe using the distance ladder disagrees with the determination using early Universe probes by ~ 4 standard deviations [19], a serious anomaly in an otherwise concordant cosmological model of our Universe, called the ΛCDM model.

0.3 The constituents of the Universe

The enormous effort put in by numerous scientists over the past 100 years in observations of the Universe and interpretation of those observations has resulted in the present era of *precision cosmology*. The precision observations of chemical elements in the oldest stars, distribution of stars and gas in the galaxies, distribution of galaxies in the Universe and study of diffuse backgrounds from microwaves to gamma rays has revealed a Universe that is remarkably simple in that the ΛCDM model with just six free parameters can explain almost all of this very diverse data.

However, the simplicity of the ΛCDM model is an illusion at best, since while assuming standard model of particle physics and general relativity as description of nature at low energies, it invokes new particles beyond the standard model of particle physics, dark matter and dark energy, to explain cosmological observations. There are already some indications that dark matter and dark energy may be more complicated than the simple 1-parameter objects of the ΛCDM model. On the positive note, cosmological observations have *discovered* new physics beyond the standard model and motivates extending the standard model of particle physics to include at least one additional new particle, the dark matter.

Dark matter - The CDM in ΛCDM stands for *cold dark matter* and it constitutes 26% of the total energy density of the Universe. The term *dark matter* or *missing mass* has been used by astronomers in the past 100 years to describe anything which was not visible with existing technology but whose presence could be inferred by the gravitational influence exerted on the visible objects in the vicinity such as stars. An interesting historical account can be found in the review by Bertone and Hooper [20].

Our solar system provides an interesting case study. In 1846 astronomers (Urbain Le Verrier in France and John Couch Adams in England) observed anomalies or deviations in the orbit of Uranus as calculated using Newton's law of gravity. They proposed that the gravitational influence of a new planet, Neptune, could explain these deviations and predicted the orbit of the new planet leading to its discovery. This is the case where a missing mass (in the form of a planet) was predicted and verified with existing technology.

Le Verrier also noticed the anomalous precession of the perihelion, the point in orbit closest to the Sun, of Mercury and proposed a similar solution: a new planet close to Sun. This time however there was no missing mass causing the anomaly. The correct solution had to wait several decades and turned out to be modification of the law of gravity - Einstein's theory of general relativity.

The first astrophysical evidences for the widespread existence of unseen matter in the Universe came from observing the motion of stars on the outskirts of galaxies (Galactic rotation curves) and motion of galaxies in galaxy clusters. The realization that most of this mass should come from an entirely new particle, not present in the standard model of particle physics, took several decades as one by one all other alternatives were discarded. In particular, the new technologies made radio observations possible which ruled out cold gas to be most of the missing mass. Advances in space exploration and technology made possible X-ray observations of astronomical objects which ruled out hot ionized gas to be the dominant component of missing mass or dark matter. Although it turned out that most of the baryonic matter was in the form of hot X-ray emitting plasma in-between the galaxies rather than in the galaxies.

Precision observations of the cosmic microwave background (CMB) and the distribution of galaxies in the Universe as well as dynamics of galaxies inside the clusters of galaxies and dynamics of stars and gas inside the galaxies has put strong limits on the interaction between ordinary (standard model) matter and dark matter. The observations also point strongly against the modification of gravity theory (general relativity) as a possible solution to the missing mass problem.

We are today at a point where we can declare quite confidently that we have indeed discovered a new particle, with non-zero mass, that is not in the standard model of particle physics. Precise astrophysical and cosmological observations as well as lab experiments put strong upper limits on the interactions this new particle (or particles) can have with the standard model particles and with itself. Apart from this we know very little about the *dark sector* of particle physics as there has been no conclusive *direct* or *indirect* detection of this particle. There are a plethora of possibilities that have been proposed of what the new particle can be.

Dark energy - The Λ in ΛCDM refers to the Einstein's cosmological constant. The cosmological constant is the simplest form of *dark energy*, the latter a generic term refering to the observation that the expansion of the Universe is accelerating. About $\approx 69\%$ of the total energy density of the Universe is dark energy today. This is contrary to the expectation from the theory of general relativity as applied to *normal* matter and radiation which have positive energy density and positive pressure. Since gravity is always an attractive force, we expect that for a Universe filled with only normal matter and radiation (including dark matter) the expansion rate will slow down with time. This implies that dark energy must be a completely different kind of fluid and in particular should have *negative* pressure. The vaccum in quantum field theory has exactly the required property, however any attempts to calculate the energy density of vaccum throw numbers hundreds of orders of magnitude larger than what is observed. A second solution, perhaps a little unsatisfactory, is that the dark energy is just a new parameter of nature, the cosmological constant. Most of the theoretical research effort has been towards getting third alternative, a dynamical fluid which behaves similar to but not exactly as the cosmological constant. The current observational effort is concentrated on finding deviations from a cosmological constant which would indicate a dynamical dark energy i.e. another new particle or physics beyond the standard model.

Inflaton - Inflaton refers to a dynamical dark energy like fluid but one which exists not today but in the very early Universe and drove an initial phase of accelerated expansion in the Universe. Part of attractiveness of the inflation theory lies in the fact that a single mechanism gives us not only an almost homogeneous and isotropic Universe but also seeds tiny initial fluctuations with a power spectrum that is naturally close to the one that is observed in the correlations of temperature fluctuations of the CMB and distribution and clustering of galaxies.

Photons - The cosmic microwave background, the oldest light in the Universe, consitutes most of the photons in the Universe by number. The detailed study of the CMB and information gathered from it forms the foundation for the precision cosmology and the standard ΛCDM model of cosmology. Once stars and galaxies have formed, they also radiate in radio, optical X-rays and gamma rays, with their combined light forming the cosmic backgrounds in these electromagnetic bands. In principle, therefore the light in any other electromagnetic band can be resolved into galaxies with sufficiently powerful telescope. The CMB is however a truly diffuse background.

Neutrinos - The massive neutrinos influence the CMB and matter distribution in the Universe in subtle ways. They are also important and participate in the nuclear reactions in the first 3 minutes of the Universe after the *big bang* which created the first light elements.

Baryons - The electrons, ions and atoms, out of which everything we see around us is made of, are collectively termed baryons in cosmology. This is the component that we can directly see and measure with photons although they form only a small fraction, $\approx 5\%$, of total energy density of the Universe.

Chapter 1

The expanding Universe

Recall: Modern cosmology began with two fundamental discoveries about our Universe in the 1920s.

1. The Universe is big

2. The Universe is expanding

Before these discoveries our Galaxy, the Milky Way, was the whole Universe, although there were speculations about other *island universes* like the Milky Way. Nevertheless, philosophers had speculated about an infinite and eternal Universe filled with stars. This cosmological model however lead to a paradox or an anomaly known as Olber's paradox, named after Olber who formulated it in modern scientific terms.

1.1 Conservation of surface brightness and Olber's paradox

We have already encountered luminosity, L, which is energy emitted by an object per unit time and flux F, which is the energy received by a detector per unit time and per unit area. In general we are interested in the spectrum of photons that are being emitted or received, i.e. energy per unit frequency. We will denote the quantities which are functions of frequency, usually measured per unit frequency, with a subscript ν. Thus,

$$I_\nu \equiv \frac{\mathrm{d}E}{\mathrm{d}\nu \mathrm{d}t \mathrm{d}A \mathrm{d}\Omega} = \frac{\text{Energy}}{\text{frequency. time. normal area. solid angle}}, \tag{1.1}$$
$$F_\nu \equiv \frac{\mathrm{d}E}{\mathrm{d}\nu \mathrm{d}t \mathrm{d}A} = \int \mathrm{d}\Omega \cos\theta I_\nu$$
$$L_\nu \equiv \frac{\mathrm{d}E}{\mathrm{d}\nu \mathrm{d}t} = \int \mathrm{d}A F_\nu$$

where E is the energy , ν is the frequency, t is time, A is effective surface area *normal* to the light rays of emitting object of absorbing detector and Ω is the solid angle. We have also defined intensity or surface brightness, I_ν. If the detector is oriented such that there is an angle θ between the direction of light rays and normal to the detector area, then the flux crossing the effective area is lowered by an amount $\cos\theta$.

An important property of the intensity, I_ν, is that it is conserved along the light rays (in non-expanding non-curved space-time). Lets have two extended sources, S1 and S2, which are at different distances d_1 and d_2 from the observer. The sources are extended and in particular have an angular size much larger than the angular resolution of the detector so that they are resolved, as shown in Fig. 1.1. Let the two sources have the same temperature (T) and therefore have the same surface brightness and emit the same blackbody spectrum with intensity

$$I_\nu = B_\nu = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/(k_\mathrm{B}T)} - 1}, \tag{1.2}$$

where h is Planck's constant, k_B is the Boltzmann constant, and c is the speed of light. The energy per unit time per unit frequency received by the observer in solid angle $\mathrm{d}\Omega$ and detector of size $\mathrm{d}A$ oriented orthogonal to the direction of the source is equal to the energy emitted by the area $\mathrm{d}A_1$ of the source in solid angle $\mathrm{d}\Omega_1$

$$\frac{\mathrm{d}E_1}{\mathrm{d}\nu \mathrm{d}t} = I_\nu \mathrm{d}\Omega_1 \mathrm{d}A_1. \tag{1.3}$$

We also have the relations $\mathrm{d}\Omega_1 = \mathrm{d}A/d_1^2$ and $\mathrm{d}\Omega = \mathrm{d}A_1/d_1^2$, where d_1 is the distance between S1 and the observer. We therefore have for the intensity $I_{\nu 1}$ measured by the observer from S1,

$$\begin{aligned} I_{\nu 1} &= \frac{\mathrm{d}E_1}{\mathrm{d}\nu \mathrm{d}t \mathrm{d}A \mathrm{d}\Omega} = I_\nu \frac{\mathrm{d}\Omega_1 \mathrm{d}A_1}{\mathrm{d}A \mathrm{d}\Omega} \\ &= I_\nu \\ &= I_{\nu 2}, \end{aligned} \tag{1.4}$$

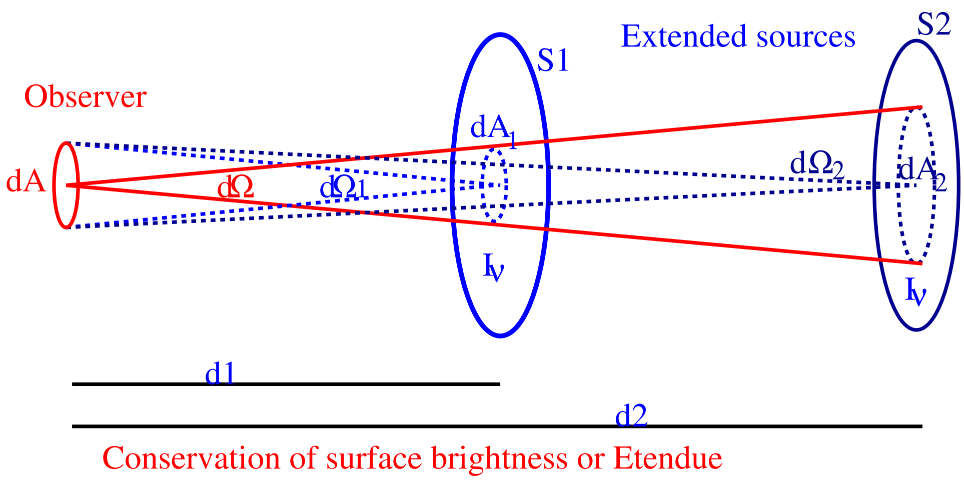

Figure 1.1

where $I_{\nu 2}$ is the intensity measured by the observer from source S2 and the last line follows from repeating the calculation for source S2. The quantity $\mathrm{d}\Omega_1 \mathrm{d}A_1$ is called the etendue and we have

$$\mathrm{d}\Omega_1 \mathrm{d}A_1 = \frac{\mathrm{d}A}{d_1^2}.\mathrm{d}\Omega d_1^2 = \mathrm{d}A.\mathrm{d}\Omega = \mathrm{d}\Omega_2 \mathrm{d}A_2. \quad (1.5)$$

The etendue and therefore intensity I_ν are conserved along the light rays.

Olber's paradox - In an infinite perpetual Universe every line of sight must eventually end on the surface of a star. Conservation of surface brightness then impliesn that we should measure the same intensity as that from the surface of a star, like the sun, along every line of sight. Therefore the entire sky should appear as bright as Sun even during the night. Why does then the night sky appear dark ? This paradox directly implies that the cosmological model that was assumed must be wrong.

Exercise 4

How does the current cosmological model of *big bang cosmology* resolve Olber's paradox ?

1.2 Newtonian cosmology

In order to have a consistent description of Universe on very large scales, we need a relativistic theory of gravity, the Einstein's theory of general relativity. However, if we confine ourselves to scales such that the light crossing time is much shorter than any other important dynamical times scale, Newtonian gravity is a valid description. We can get useful insights into the expanding Universe using Newtonian gravity, before moving to a relativistic description.

Let us consider a dust sphere of uniform density $\rho(t)$ uniformly expanding with time t. The mass shell at radius $R(t)$ has velocity $\upsilon(t)$ directed away from the center of the sphere in the radial direction. Since we have spherical symmetry, according to Gauss's theorem, the gravitational force on a shell at radius $(R(t)$ only depends on the mass inside the shell and we can assume that all the mass is concentrated at the center to calculate the gravitational force on the shell. Also, we will assume uniform expansion, which is enforced by choosing suitable initial conditions. In a uniformly expanding homogeneous sphere the different shells do not cross, so mass inside a shell at $R(t)$ is independent of time. The acceleration of the shell is given by

$$\ddot{R}(t) = -\frac{GM\left(R(t)\right)}{R(t)^2}, \tag{1.6}$$

where an over-dot represents derivative w.r.t. time t. Using $M\left(R(t)\right) = 4/3\pi R(t)^3\rho(t)$ we can rewrite the above equation as

$$\frac{\ddot{R}}{R} = -\frac{4\pi}{3}G\rho(t) \tag{1.7}$$

This is Friedmann's second equation. We get exactly the same equation in general relativity for a dust Universe, i.e. for a Universe made of matter such that the pressure (P) is negligible compared to the energy density, $P \ll \rho$. If the pressure cannot be neglected there is an additional term which we cannot get in a Newtonian calculation since in Newtonian gravity only the energy density is the source of gravity while in Einsteinian gravity or general relativity the stress energy tensor, which includes pressure, is the source of gravity.

Since the energy density is always positive, $\rho > 0$, a Newtonian Universe can never accelerate in the expanding phase ($\dot{R} > 0$) and must always decelerate, i.e. $\ddot{R} < 0$. In particular, there is no dark energy in a Newtonian Universe. Whether this deceleration will eventually stop the Universe from expanding and cause the Universe to collapse depends on the initial conditions. To be more precise it depends on whether the total energy of the Universe is positive, zero, or negative.

Lets consider a thin mass shell at distance $R(t)$ from the center with mass $\mathrm{d}m = 4\pi R^2\mathrm{d}R\,\rho(t)$. The total energy C' of this mass shell is given by the sum of kinetic energy (T) and gravitational potential energy (V),

$$
\begin{aligned}
C' &= T + V \\
&= \frac{1}{2}\mathrm{d}m\dot{R}^2 - \frac{GM(R)\mathrm{d}m}{R}
\end{aligned}
\tag{1.8}
$$

The mass of the shell dm is constant as we follow this shell and the Newtonian sphere expands. Defining $C = 2C'/\mathrm{d}m$, we can rewrite the above equation of energy conservation as

$$
\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} + \frac{C}{R^2} \tag{1.9}
$$

This is Friedmann's first equation or just Friedmann equation.

The constant C is the total energy per unit mass and is a conserved quantity set by initial conditions.

- $\mathbf{C < 0}$: Closed or bounded Universe. The dust sphere will reach maximum expansion and collapse back.
- $\mathbf{C = 0}$: Flat or just unbounded Universe. The dust sphere will expand forever with velocities of mass shells approaching zero as $t \to \infty$.
- $\mathbf{C > 0}$: Open or unbounded Universe. Expansion velocity remains finite as $t \to \infty$.

Analogy - Throwing a ball upward on earth with velocity smaller, equal to, or greater than the escape velocity.

The interpretation in terms of closed, flat or open geometry corresponding to positive, zero, or negtaive curvature can only be done in general relativity with the constant C identified as curvature of space. Note that there is also no cosmological constant in Newtonian cosmology.

In order to have a uniform expanding sphere, we have to fine tune the initial conditions very finely in Newtonian cosmology. In particular, in a uniformly expanding Universe the different mass shells do not cross. We therefore need that initially the outer shells should have larger velocities compared to the inner shells closer to the center, i.e. the sphere should follow Hubble's law, $v \propto R$. This is most easily seen by looking at the bounded Universe. If we have the initial conditions just right (Hubble's law), outer shells which started with higher initial velocities compared to the inner shells also suffer larger deceleration and all shells will come to rest at maximum expansion at the same time.

Lets assume that we have the needed finely tuned initial conditions. Then mass conservation means that mass inside a shell at radius R,

$$M(R) = \text{constant} = \frac{4\pi}{3}R^3\rho$$
$$\rho \propto R^{-3} = \rho_0\left(\frac{R}{R_0}\right)^{-3}, \tag{1.10}$$

where R_0 is some reference radius at a time t_0 when the density is ρ_0.

The Friedmann equation can be written as

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho_0}{3}\left(\frac{R}{R_0}\right)^{-3} + \frac{C}{R^2} \tag{1.11}$$

We can now solve this for R as function of time.

$$\dot{R}^2 = \frac{8\pi G\rho_0 R_0^3}{3R} + C$$
$$= \frac{2GM}{R} + C$$
$$\dot{R} = \pm\left(\frac{2GM}{R} + C\right)^{1/2}, \tag{1.12}$$

where the + sign corresponds to the expanding Universe and − sign to the contracting Universe.

This equation is easily solved for a flat Universe, i.e. when $C = 0$, giving

$$R = (2GM)^{1/2}\, t^{2/3} + \text{constant}. \tag{1.13}$$

The initial conditions $R = 0$ at $t = 0$ set the constant of integration to 0.

For the general case of $C \neq 0$, we can solve by doing a change of variables from time t to a new variable η defined by $\mathrm{d}t = R\mathrm{d}\eta$ so that

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{1}{R}\frac{\mathrm{d}}{\mathrm{d}\eta}. \tag{1.14}$$

The Eq. 1.12 become in terms of new variables,

$$\frac{\mathrm{d}R}{\mathrm{d}\eta} = \left(2GMR + CR^2\right)^{1/2} \tag{1.15}$$

with the solution

$$R(\eta) = \frac{GM}{C}\left[\frac{e^{-\sqrt{C}\eta} + e^{\sqrt{C}\eta}}{2} - 1\right] \tag{1.16}$$

Thus for $C > 0$ we have

$$R(\eta) = \frac{GM}{C}\left[\cosh\left(\sqrt{C}\eta\right) - 1\right] \tag{1.17}$$

and for $C < 0$,

$$R(\eta) = \frac{GM}{|C|}\left[1 - \cos\left(\sqrt{|C|}\eta\right)\right] \tag{1.18}$$

We can also integrate $\mathrm{d}t = R\mathrm{d}\eta$ to get, for $C > 0$,

$$t = \frac{GM}{C^{3/2}}\left[\sinh\left(\sqrt{C}\eta\right) - \eta\right] \tag{1.19}$$

and for $C < 0$

$$t = \frac{GM}{|C|^{3/2}}\left[\eta - \sin\left(\sqrt{|C|}\eta\right)\right]. \tag{1.20}$$

We thus have a parameteric solution for the Friedmann equation for non-flat universes.

Exercise 5

Shell crossing: Lets consider two shells infinitesimal distance apart at time $t = t_0$ with one shell with radius $R_1(t_0) = R$ and second shell at $R_2(t_0) = R + \mathrm{d}R$, with $R_1 < R_2$ at t_0. As long as the shells do not cross, the mass inside each of the two shells remains constant and the solutions we have derived are valid. For the shells to cross at some later time t^*, they must coincide, i.e. $R_1(t^*) = R_2(t^*)$. Show that this can only happen at the origin when $R_1(t^*) = R_2(t^*) = 0$, i.e. only as a initial condition for flat and open Universes and additionally also at the time of *big crunch* for closed universes. **Bonus:** The shell crossing is avoided because of our fine tuning of the parameters of the Universe. Identify where and how this fine tuning is happening.

1.3 Relativistic cosmology and the Friedmann metric

There are several limitations of the Newtonian world model. Since Newton's theory is a good approximation only on small scales, Newtonian Universe must necessarily be finite. An infinite Newtonian Universe does not make much sense. We can apply Newton's law to small parts of the Universe but not the Universe as a whole. To study the Universe as a whole we need a relativistic theory of gravity. Even though the equations for the evolution of spherical shells in Newtonian world model look similar to the Friedmann's equations in general relativity, the meaning of these equations and in particular the *expansion of the Universe* is completely different in Newtonian gravity and general relativity. In Newtonian Universe, spacetime is fixed and expansion is just matter moving through it. However, in such a case, uniform expansion requires extremely fine tuned initial conditions, i.e. different mass shells must follow Hubble's law exactly. Hubble's law is however not an initial condition in Friedmann Universe but a prediction coming from the application of Einstein's equation of general relativity to a homogeneous and isotropic Universe. Friedmann applied the equations of general relativity to the whole Universe soon after the theory was proposed by Einstein. Friedmann found three different homogeneous and isotropic solutions for the Universe which differed only in the value of the spatial curvature of the Universe i.e. flat, closed and open universes [15, 16].

The simplest or the most symmetric Universe we can conceive of would be the one that is homogeneous and isotropic. Put another way, in such a Universe there exists a coordinate system $(t, \mathbf{x})$, where t is proper time and $\mathbf{x}$ is the three dimensional spatial coordinates vector, such that at any time t the Universe is homogeneous and istotropic or the matter is distributed in a homogeneous and isotropic manner. In particular the density and pressure of matter in such a Universe (or more precisely the stress energy tensor $T^{\mu\nu}$) are only a function of time but independent of the spatial coordinates $\mathbf{x}$. The question that we want to answer is: what is the metric $g_{\mu\nu}$ or the line element

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu, \tag{1.21}$$

which satisfies Einstein's equations for such a stress energy tensor, where summation is implied over repeated indices with greek indices running from 0 to 3, 0 being the time coordinate, and $\mu = 1 - 3$ the spatial coordinates. Note that we can always do a change in the coordinate system so that in the new coordinates the Universe no longer *looks* homogeneous and isotropic, i.e. $g_{\mu\nu}$ or $T^{\mu\nu}$ is a function of spatial coordinates. One example is starting with coordinates in which the Universe is homogeneous and isotropic and giving a boost to some or all observers. In a isotropic Universe the CMB will have same temperature in all directions. In fact, the CMB that we observe from Earth is not isotropic. It has a large dipole with slightly large temperature in one direction compared to the opposite direction. However, this is simply because observers on Earth (or on a satellite going around earth) are not *comoving* observers i.e. observers at rest in the reference frame defined by a homogeneous and isotropic Friedmann metric, but have a *peculiar velocity*. Earth is moving around the Sun at $\sim$ 30 km/s, Sun is going around the galaxy at $\sim$ 220 km/s and our galaxy is falling towards local concentration of galaxies with the combination of all these motions giving us a velocity of $\sim$ 300 km/s w.r.t. the

CMB. However, observation of this dipole in the CMB does not make the Universe anisotropic. We can simply do a Lorentz boost, undoing the effect of our net motion w.r.t the CMB and recover isotropic CMB. This is just the freedom to choose coordinates given by general relativity and universes related by *smooth* coordinate transformations are identified to be the same. We would always choose coordinates in which the metric and stress energy tensor look simplest.

A homogeneous and isotropic Universe satisfies the *cosmological principle* trivially. The cosmological principle is an assumption about our place in the Universe, the Copernical principle applied to the Universe as a whole: We are not in a special place in the Universe. The finite Newtonian Universe does not satisfy the cosmological principle since it has a definite centre. The Friedmann Universe does not have any centre, all observers are equivalent and in particular the Universe looks identical to all observers. We observe the cosmic microwave background, the oldest light in the Universe, and it appears to us as isotropic. The Copernican principle states that we are not in any special place in the Universe, therefore CMB must appear as isotropic to every observer in the Universe. Thus the Universe must be homogeneous (on large scales) in addition to being isotropic to us.

Exercise 6

Prove that a Universe which is isotropic about every point in space is also homogeneous.

Metrics with curvature - The metric is a purely geometric quantity which tells us how to measure distances in any space and exists independent of the Einstein's equations of general relativity. Lets consider a 2-dimensional sphere or 2-sphere, $\mathbb{S}^2$. This is just the 2-dimensional surface of an ordinary sphere in 3-dimensional flat space, $\mathbb{R}^3$. The metric on $\mathbb{R}^3$ is the usual Euclidean metric in flat space,

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2, \tag{1.22}$$

where x, y, z are the coordinates. The sphere is defined by the condition (choosing origin of sphere at the origin of coordinate system)

$$x^2 + y^2 + z^2 = \text{constant} = \frac{1}{K'}$$

However, we are describing 2-D object with 3 numbers. Ideally we want to choose 2-D coordinates on 2-D objects. This can be accomplished by eliminating one of the coordinates using the above constraint. Taking the derivative we get,

$$x\mathrm{d}x + y\mathrm{d}y + z\mathrm{d}z = 0$$
$$\mathrm{d}z = -\frac{x\mathrm{d}x + y\mathrm{d}y}{\left[1/K' - (x^2 + y^2)\right]^{1/2}}. \tag{1.23}$$

Thus we can rewrite the *metric on the sphere*, i.e. how we measure distances between two points on the sphere along any curve that remains on the sphere, in terms of just two coordinates,

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \frac{K'\left(x\mathrm{d}x + y\mathrm{d}y\right)^2}{1 - K'\left(x^2 + y^2\right)} \tag{1.24}$$

This is the metric inherited by the 2-sphere from the 3-D space it is embedded in. In general, nothing is stopping us from having a different definition of distance on the sphere and choosing a quite different metric. The metric in Eq. 1.24 is an explicitly 2-D metric, depending only on 2 coordinates, x and y, and tells us how to measure distances *locally*. Note that at the origin of the coordinates, $x = y = 0$, the metric reduces to a flat or Euclidean metric. Away from the origin, the metric departs from a Euclidean one. We can do a change of coordinates to spherical polar coordinates,

$$\begin{aligned} x = r\cos\theta,\ y = r\sin\theta \\ \mathrm{d}x^2 + \mathrm{d}y^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 \\ x^2 + y^2 = r^2, x\mathrm{d}x + y\mathrm{d}y = r\mathrm{d}r \end{aligned} \tag{1.25}$$

The metric on the 2-sphere in terms of new coordinates becomes,

$$\begin{aligned} \mathrm{d}s^2 &= \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + \frac{K' r^2 \mathrm{d}r^2}{1 - K' r^2} \\ &= \frac{\mathrm{d}r^2}{1 - K' r^2} + r^2\mathrm{d}\theta^2. \end{aligned} \tag{1.26}$$

We note that this coordinate system does not cover the whole sphere. In particular there is a coordinate singularity at $r = \sqrt{1/K'}$, it is not a real singularity and can be removed by a coordinate transformation. We will come back to this point.

Similarly, we can look at 3-sphere, $\mathbb{S}^3$, which is more relevant for cosmology, embedded in a 4-D Euclidean space, $\mathbb{R}^4$ with metric,

$$\begin{aligned} \mathrm{d}s^2 &= \left(\mathrm{d}x^1\right)^2 + \left(\mathrm{d}x^2\right)^2 + \left(\mathrm{d}x^3\right)^2 + \left(\mathrm{d}x^4\right)^2 \\ &= \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j, \end{aligned} \tag{1.27}$$

where δ_{ij} is the Kronecker delta function with the definition $\delta_{ij} = 1$ for $i = j$ and 0 otherwise. The 3-sphere is defined by

$$\delta_{ij}x^i x^j = \frac{1}{K'},\ \ i, j = 1 \text{ to } 4 \tag{1.28}$$

As before, we can write the 4th coordinate, x^4 in terms of the other three.

$$\delta_{ij}x^i\mathrm{d}x^j = 0$$

$$\mathrm{d}x^4 = \frac{-K'^{1/2}\left(x_1\mathrm{d}x^1 + x_2\mathrm{d}x^2 + x_3\mathrm{d}x^3\right)}{\left[1 - K'\left(x_1^2 + x_2^2 + x_3^2\right)\right]^{1/2}} \tag{1.29}$$

giving the metric on $\mathbb{S}^3$,

$$\mathrm{d}s^2 = \left(\mathrm{d}x^1\right)^2 + \left(\mathrm{d}x^2\right)^2 + \left(\mathrm{d}x^3\right)^2 + \frac{K'\left(x_1\mathrm{d}x^1 + x_2\mathrm{d}x^2 + x_3\mathrm{d}x^3\right)^2}{1 - K'\left(x_1^2 + x_2^2 + x_3^2\right)} \tag{1.30}$$

We can again do a coordinate transformation to spherical polar coordinates,

$$\begin{aligned} x_1 &= r\cos\theta\sin\phi \\ x_2 &= r\cos\theta\cos\phi \\ x_3 &= r\sin\theta \end{aligned} \tag{1.31}$$

Exercise 7

Do the conversion from Euclidean to spherical polar coordinates on $\mathbb{S}^3$.
The result is

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1 - K'r^2} + r^2\mathrm{d}\Omega,\ \mathrm{d}\Omega = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2 \tag{1.32}$$

We again see the same coordinate singularity at $r = 1/\sqrt{K'}$, i.e. when the 3rd Euclidean coordinate (z) for $\mathbb{S}^2$ and 4th coordinate for $\mathbb{S}^3$ (x_4) is zero.

Lets put the metric in standard form by rescaling the variables by a constant *scale factor* a,

$$x_i \to ax_i,\ r \to ar. \tag{1.33}$$

The metric in rescaled variables becomes

$$\mathrm{d}s^2 = a^2\left[\left(\mathrm{d}x^1\right)^2 + \left(\mathrm{d}x^2\right)^2 + \left(\mathrm{d}x^3\right)^2 + \frac{K'a^2\left(x_1\mathrm{d}x^1 + x_2\mathrm{d}x^2 + x_3\mathrm{d}x^3\right)^2}{1 - K'a^2\left(x_1^2 + x_2^2 + x_3^2\right)}\right] \tag{1.34}$$

We can additionaly choose $a = 1/\sqrt{K'}$, $K = K'a^2 = 1$ giving,

$$\mathrm{d}s^2 = a^2\left[\left(\mathrm{d}x^1\right)^2 + \left(\mathrm{d}x^2\right)^2 + \left(\mathrm{d}x^3\right)^2 + \frac{K\left(x_1\mathrm{d}x^1 + x_2\mathrm{d}x^2 + x_3\mathrm{d}x^3\right)^2}{1 - K\left(x_1^2 + x_2^2 + x_3^2\right)}\right], \tag{1.35}$$

with $K = +1$, i.e. we have scaled out the dimensions of our sphere, and a is now the *radius of curvature* or the size of the closed Universe, the sphere $\mathbb{S}^3$. In spherical coordinates we get

$$\mathrm{d}s^2 = a^2\left[\frac{\mathrm{d}r^2}{1-Kr^2} + r^2\mathrm{d}\Omega\right] \tag{1.36}$$

The scale factor in a closed Universe is thus not arbitrary but is the actual size of the Universe with units of distance. The comoving coordinates, x^i or r, θ, ϕ, are defined on a sphere of size 1 or the unit sphere. We note that this is a space of constant curvature, $K = +1$ and is thus homogeneous and isotropic. Any point on the sphere is related to any other point by translation on $\mathbb{S}^3$ or equivalently rotation in $\mathbb{R}^4$. For K=0 we recover the flat or Euclidean space. The flat space is inifinte (i.e. the size of the Universe $1/K \to \infty$ and the scale factor in this case is not related to K but is arbitrary. It is conventional to normalize the scale factor for flat space so that $a = 1$ today.

What about $K = -1$? Suppose we started with a Lorentzian metric (also called semi-Euclidean metric) which is obtained from Euclidean metric by flipping the sign of the fourth coordinate,

$$\mathrm{d}s^2 = \left(\mathrm{d}x^1\right)^2 + \left(\mathrm{d}x^2\right)^2 + \left(\mathrm{d}x^3\right)^2 - \left(\mathrm{d}x^4\right)^2 \tag{1.37}$$

and looked at the *sphere* with negative radius,

$$x_1^2 + x_2^2 + x_3^2 - x_4^2 = \frac{1}{K'},\ K' < 0. \tag{1.38}$$

This is the equation for a 3-D hyperboloid embedded in a 4-d pesudo-Euclidean space (or semi-Riemannian space). A space is called Riemannian if distances are always positive and semi-Riemannian if distances can also be negative or zero. Minkowski space is semi-Riemannian. Note that the hyperboloid, even though it is embedded in a semi-Riemannian space, is itself Riemannian, i.e. distances on the hyperboloid are always positive. Following the same procedure as for the sphere $\mathbb{S}^3$, we get the metric on the hyperboloid by eliminating one of the redundant coordinates,

$$\mathrm{d}s^2 = a^2\left[\left(\mathrm{d}x^1\right)^2 + \left(\mathrm{d}x^2\right)^2 + \left(\mathrm{d}x^3\right)^2 + \frac{K\left(x_1\mathrm{d}x^1 + x_2\mathrm{d}x^2 + x_3\mathrm{d}x^3\right)^2}{1-K\left(x_1^2+x_2^2+x_3^2\right)}\right], \tag{1.39}$$

and in spherical coordinates,

$$\mathrm{d}s^2 = a^2\left[\frac{\mathrm{d}r^2}{1-Kr^2} + r^2\mathrm{d}\Omega\right] \tag{1.40}$$

with $K = -1$, $a = 1/\sqrt{-K'}$, the radius of curvature. Thus the metric has exactly the same form for open, flat and closed universes.

We can prove that these are the only possible metrics that describe homogeneous and isotropic space [21–23]. The coordinate singularities arise only in the positive curvature space, since it is topologically distinct from a Euclidean space. There is no continuous deformation that can make a sphere into Euclidean space. We are trying to wrap a Euclidean space around a sphere by using a metric inherited from a higher dimensional Euclidean space. There is no such problem in space with negative curvature. We can get rid of the cooordinate singularity for $\mathbb{S}^3$ by doing a coordinate transformation. In particular we want to use coordinates in which the *compactness* of $\mathbb{S}^3$ is built-in, i.e. if we go around a sphere we come back to the same point and the coordinates should also reflect this.

Lets start with a 1-D sphere, $\mathbb{S}^1$, i.e. a circle. An obvious choice of coordinates would be the angle θ with $\theta = 0$ and $\theta = 2\pi$ identified as the same point. The distance covered when travelling an infinitesimal distance $\mathrm{d}\theta$ is given by the metric $\mathrm{d}s^2 = a^2\mathrm{d}\theta^2$, where a is the radius of the circle. For $\mathbb{S}^2$, we can reach any point on the unit sphere using two angles, the polar angle $0 \leq \theta \leq \pi$ and the azimuthal angle $0 \leq \phi \leq 2\pi$. If we keep ϕ constant, the length of the arc traced by small change in angle θ is just $a\mathrm{d}\theta$, where a is again the actual size of the sphere. If we keep θ constant and vary ϕ, we trace a circle of radius $a\sin\theta$ and the arc length is $a\sin\theta\mathrm{d}\phi$. Thus the singularity free metric for $\mathbb{S}^2$ is

$$\mathrm{d}s^2 = a^2\left[\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right] \tag{1.41}$$

This is just the 2-d angular part of the Euclidean metric for $\mathbb{R}^3$ in spherical polar coordinates. This makes sense, since the radial coordinate is the 3rd dimension which is not needed to tell us where we are on the 2-d surface of the sphere $\mathbb{S}^2$. We need only the angular part of the 3-d Euclidean metric to reach any point on the $\mathbb{S}^2$. Similarly, for $\mathbb{S}^3$, we can choose coordinates which are the angular part of the spherical coordinates on $\mathbf{R}^4$. Alternaticely, we can directly do a coordinate transformation on the metric, Eq. 1.36, defining

$$\begin{aligned} r &= \sin\Psi,\ \mathrm{d}r = \cos\Psi\mathrm{d}\Psi,\ 1-r^2 = \cos^2\Psi \\ \mathrm{d}s^2 &= a^2\left[\mathrm{d}\Psi^2 + \sin^2\Psi\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right]. \end{aligned} \tag{1.42}$$

Note that Ψ is now an angular coordinate. After the transformation, there is no singularity in the new coordinates and we can cover the whole space.

Exercise 8

On $\mathbf{S}^2$, we can reach every point by allowing $0 \leq \theta \leq \pi$, $0 \leq \phi \leq 2\pi$. This can be see by realizing that on a unit sphere embedded in $\mathbb{R}^3$, $x = \sin\theta\cos\phi$,

$y = \sin\theta\sin\phi$, $z = \cos\theta$. The point $\theta = \pi + \alpha, \phi = \phi_0$, for any α, ϕ_0 is the same point as $\theta = \pi - \alpha, \phi = \phi_0 + \pi$. Thus there is no need for θ to go beyond π. What is the minimum range of the three angles Ψ, θ, ϕ on $\mathbb{S}^3$ that allows us to reach every point on $\mathbb{S}^3$.

For $K = -1$, we can do a coordinate change $r = \sinh\Psi$ giving the metric

$$\mathrm{d}s^2 = a^2\left[\mathrm{d}\Psi^2 + \sinh^2\Psi\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right]. \tag{1.43}$$

What does this mean? For two points separated by a small angle $\mathrm{d}\Omega$ in the θ, ϕ part of the metric, the points fly apart exponentially as the radial coordinate Ψ increases. In flat space, $K = 0$,

$$\mathrm{d}s^2 = a^2\left[\mathrm{d}\Psi^2 + \Psi^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right], \tag{1.44}$$

the distance only increases as $\propto \Psi$. The scale factor a in flat space-time is arbitrary since there is no curvature scale to define it.

We promote the homogeneous and isotropic space metric to a space-time metric by making the scale factor a a function of time, t, obtaining the Friedmann metric, where $f(\Psi)^2 = \sin^2\Psi, \Psi^2, \sinh^2\Psi$ for $K = +1, 0, -1$ respectively. It is usually convenient to define conformal time, η, as $\mathrm{d}t = a\mathrm{d}\eta$, bringing time and space to an as equal a treatment as possible. In terms of conformal time, the metric is

$$\mathrm{d}s^2 = a(\eta)^2\left[-\mathrm{d}\eta^2 + \mathrm{d}\Psi^2 + f(\Psi)^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right]. \tag{1.45}$$

Which coordinates to choose usually depends on the problem we want to solve and the aspects of physics we are interested in. We can even tolerate coordinate singularities in practice as long as we work in part of the space away from the singularities. There are strong limits on the curvature of Universe from cosmological data which imply that our observable Universe is much smaller compared to any possible curvature of the Universe. Thus for Universe at least, the coordinate singularities in Eq. 1.36 is not usually a problem.

Exercise 9

Geodesic Equation in curved space

In curved space with generalised coordinates $\mathbf{q}$, the Lagrangian of a free particle can be written as

$$\begin{aligned} L &= \frac{1}{2}mv^2 = \frac{1}{2}m\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 \\ &= \sum_{i,j}\frac{1}{2}mg_{ij}\left(\frac{\mathrm{d}q_i}{\mathrm{d}t}\frac{\mathrm{d}q_j}{\mathrm{d}t}\right), \end{aligned} \tag{1.46}$$

where $ds^2 = \sum_{ij} g_{ij} dq_i dq_j$ is the generaized length element along the geodesic of the particle and g_{ij} is the metric which in general would be a function of the generalised coordinates, $g_{ij} = g_{ij}(q)$. Show that the Euler-Lagrange equations in general curved coordinates (or curved space) are

$$\frac{d^2 q_i}{dt^2} + \sum_{jk} \Gamma_{ijk} \frac{dq_j}{dt}\frac{dq_k}{dt} = 0,$$
$$\text{where } \Gamma_{ijk} = \sum_{\ell} \frac{1}{2}\tilde{g}_{i\ell}\left[\frac{\partial g_{k\ell}}{\partial q_j} + \frac{\partial g_{\ell j}}{\partial q_k} - \frac{\partial g_{jk}}{\partial q_\ell}\right], \tag{1.47}$$

$\tilde{g}_{ij}$ is the inverse metric defined by the equations $\tilde{g}_{ij} g_{jk} = \delta_{ik}$, δ_{ik} is the Kronecker delta with $\delta_{ik} = 1$ if $i = k$ and zero otherwise i.e. δ_{ij} is the identity matrix and the matrix $\tilde{g}$ is the matrix inverse of g.

1.4 Friedmann Equations

Friedmann showed that the homogeneous and isotropic metric, Eq. 1.45, is a solution to the Einstein's equations of general relativity. What this means is that there exists a stress energy tensor, $T^{\mu\nu}$, such that the Einstein's equations are satisfied with the metric given by the Friedmann metric. This stress energy tensor is in particular also homogeneous and isotropic. In addition, general relativity also gives us the equations for the time evolution of the stress energy tensor and the scale factor, the only quantities which are a function of time. The question then arises whether our Universe is such a Universe described by the Friedmann's solutions ?

The CMB and distribution of galaxies in the Universe is, on large scales, homogeneous and isotropic. The CMB, in fact, only tells us that the Universe is isotropic from our position. The Copernican principle, that we are not in a special place in the Universe, implies isotropy about every space-time point and therefore homogeneity. With the recent large scale galaxy surveys, we have been able to test the assumption of Copernican principle and the Universe is indeed found to be homogeneous on large scales (scales $\gtrsim$ 350 Mpc [24]).

Thus observations, in addition to theoretical bias towards simplicity, motivate the choice of Friedmann metric (also known as Friedmann-Lemaitre-Robertson-Walker or FLRW metric [see 18, for historical context]) for the zeroth order cosmological solution, using spherical coordinates (r, θ, ϕ),

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu$$
$$g_{rr} = \frac{a^2}{1 - Kr^2},\ g_{\theta\theta} = a^2 r^2,\ g_{\phi\phi} = a^2 r^2 \sin^2\theta,\ g_{00} = -1 \tag{1.48}$$

Exercise 10

Find the inverse metric $g^{\mu\nu}$ defined by equation $g_{\mu\nu}g^{\nu\alpha} = \delta^{\alpha}_{\mu}$.

The inverse metric $g^{\mu\nu}$ is therefore different from the metric $g_{\mu\nu}$, and in general a tensor with upper indices is a different object (living in a dual vector space to be precise) with different physical meaning compared to the same tensor with a lower indices and we have to be careful in placing the indices. The metric tensosrs with upper and lower indices $g_{\alpha\beta}, g^{\alpha\beta}$ map a tensor to its dual space, i.e. they can be used to lower or raise the index. Thus, for a general tensor $T^{\alpha\beta\gamma...}g_{\alpha\mu} = T_{\mu}^{\beta\gamma...}$ and $T_{\alpha\beta\gamma...}g^{\beta\mu} = T^{\mu}_{\alpha\gamma...}$.

The Hubble parameter H is defined as

$$H \equiv \frac{\dot{a}}{a} \equiv \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} \tag{1.49}$$

The coordinates r, θ, ϕ are comoving coordinates. The physical distance d at constant time t between two points is given by choosing one point at the origin of the coordinate system and with the other point having the coordinate r, without loss of generality, by

$$d(r,t) = \int_0^r \mathrm{d}r\,\sqrt{g_{rr}} = a(t)\int_0^r \frac{\mathrm{d}r}{\sqrt{1-Kr^2}} = a(t)\begin{cases} \sin^{-1} r, & K=+1, r<1 \\ \sinh^{-1} r, & K=-1 \\ r, & K=0 \end{cases} \tag{1.50}$$

The comoving distance between two *comoving observers* (zero peculiar velocity or coordinate velocity) is independent of time while the physical distance scales as the scale factor. This is true for all values of global curvature (K). The proper time interval measured by the comoving observers (i.e. observers with r, θ, ϕ constant) is just $-g_{\mu\nu}\mathrm{d}x^{\mu}\mathrm{d}x^{\nu} = \mathrm{d}t^2$, t is therefore the proper time as measured by the comoving observers.

1.4.1 Geodesics in Friedmann Universe

We will use the notation that greek indices (e.g. μ, ν, α, β) refer to space-time and run from 0 to 3 while latin indices (i, j...) refer to purely spatial components of a tensor and run from 1 to 3.

Lets look at the geodesics of particles in an expanding Universe described by Friedmann metric. The geodesic equation for a particle with four-momentum

$$P^{\mu} \equiv \frac{\mathrm{d}x^{\mu}}{\mathrm{d}\lambda}, \tag{1.51}$$

where λ is an affine parameter, is

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu_{\alpha\beta}\frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\beta}{\mathrm{d}\lambda} = 0$$

or

$$\frac{\mathrm{d}P^\mu}{\mathrm{d}\lambda} + \Gamma^\mu_{\alpha\beta}P^\alpha P^\beta = 0 \tag{1.52}$$

The Christoffel symbols $\Gamma^\mu_{\alpha\beta}$ are given by

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2}g^{\mu\gamma}\left[\frac{\partial g_{\gamma\alpha}}{\partial x^\beta} + \frac{\partial g_{\gamma\beta}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\right] \tag{1.53}$$

We note that since g_{00} is constant and the off diagonal components of the metric are zero, we get for the Christoffel symbols $\Gamma^\gamma_{\alpha\beta}$,

$$\Gamma^i_{00} = 0 \tag{1.54}$$

This implies that for a particle at rest, $P^i = 0$, the geodesic equation reduces to $\mathrm{d}P^i/\mathrm{d}\lambda = 0$ and implies that a particle at rest will remain at rest i.e. the acceleration is zero.

We want to find the momentum-energy of a particle with 4-momentum P^μ as measured by a comoving observer, i.e. an observer who is at rest in the Friedmann metric with 4-velocity $V^\mu \equiv (1, \vec{0})$ and how it evolves with time. The energy measured by observer with 4-velocity V^μ is given by

$$E = -V^\mu P_\mu = -V^0 P_0 = -P_0 \tag{1.55}$$

The geodesic equation we need is therefore the one for the zeroth component of 4-momentum,

$$\frac{\mathrm{d}P^0}{\mathrm{d}\lambda} + \Gamma^0_{\alpha\beta}P^\alpha P^\beta = 0 \tag{1.56}$$

The only non-zero components of $\Gamma^0_{\alpha\beta}$ are

$$\begin{aligned}\Gamma^0_{rr} &= \frac{1}{2}g^{00}\left[-\frac{\partial g_{rr}}{\partial t}\right] = \frac{1}{2}2\dot{a}a\frac{1}{1-Kr^2} = \frac{a^2}{1-Kr^2}\frac{\dot{a}}{a}\\ &= \frac{\dot{a}}{a}g_{rr}\end{aligned} \tag{1.57}$$

Similarly for $\Gamma^0_{\theta\theta}$ and $\Gamma^0_{\phi\phi}$ giving

$$\Gamma^0_{ij} = \frac{\dot{a}}{a}g_{ij} \tag{1.58}$$

$$\frac{dP^0}{d\lambda} + \frac{\dot{a}}{a} g_{ij} P^i P^j = 0 \tag{1.59}$$

For any particle, massive or massless, the square of the 4-momentum is a constant (scalar) equal to the square of its mass,

$$-m^2 = P_\alpha P^\beta = g_{\alpha\beta} P^\alpha P^\beta = g_{00} P^0 P^0 + g_{ij} P^i P^j. \tag{1.60}$$

We identify the spatial part in above equation, $g_{ij}P^iP^j$, as the square magnitude of 3-momentum (p) measured by the observer, $g_{ij}P^iP^j \equiv p^2$ and

$$P_0 = g_{\alpha 0} P^\alpha = g_{00} P^0 = -P^0 \equiv -E, \tag{1.61}$$

where $E \equiv P^0 \equiv dt/d\lambda$ is the energy of the particle measured by the comoving observer. Therefore, Eq. 1.60 is just the GR version of relation $E^2 = p^2 + m^2$. We thus have,

$$2E\frac{dE}{d\lambda} = 2p\frac{dp}{d\lambda} \tag{1.62}$$

We can therefore write Eq. 1.59 as

$$\begin{aligned} \frac{p}{E}\frac{dp}{d\lambda} + \frac{\dot{a}}{a}p^2 &= 0 \\ \frac{dp}{dt} + \frac{\dot{a}}{a}p &= 0 \end{aligned} \tag{1.63}$$

The solution to the above equation is

$$p \propto 1/a \tag{1.64}$$

for all particles and irrespective of the global curvature K. This is an important result and tells us that the momenta of all particles *redshift* inversely proportional to the scale factor. This result has several implications.

The physical volume element of a region with edges $dr, d\theta, d\phi$ is given by, using the metric, $\Delta V = a^3 f(r,\theta,\phi,K)drd\theta d\phi$, where $f(r,\theta,\phi,K)$ is a fixed function that depends on the geometry or curvature. Since the comoving coordinates r,θ,ϕ of a region are fixed, $\Delta V \propto a(t)^3$. Also from Eq. 1.64, we conclude that the volume in momentum space, $d^3p \propto a^{-3}$. The phase space volume of a distribution of particles, d^3pd^3x is therefore conserved i.e. it is independent of the scale factor and does not change with the expansion of the Universe. If no particles are created or destroyed, the number of particles, N, is also conserved. Thus the phase space density or the occupation number, $\mathrm{d}N/(\mathrm{d}^3p\mathrm{d}^3x)$ is a conserved quantity. For photons (or any boson) with a Bose-Einstein spectrum, the occupation number f_{BE} with chemical potential μ and temperature T is given by

$$f_{\mathrm{BE}}(p) = \frac{\mathrm{d}N}{\mathrm{d}^3p\mathrm{d}^3x} = \frac{1}{e^{(p-\mu)/T} - 1} \tag{1.65}$$

is conserved, implying that $(p-\mu)/T$ is conserved. Since $p \propto 1/a$, conservation of ocupation number implies that $T, \mu \propto 1/a$. Same law also holds for fermions with a Fermi-Dirac distribution (i.e neutrinos). In particular, for the CMB with a initial Planck spectrum (i.e. Bose-Einstein spectrum with $\mu = 0$) or neutrinos with an initial Fermi-Dirac spectrum, the expansion of the Universe does not change the spectral shape and the spectrum remains Planck (blackbody) or Fermi-Dirac in the absence of momentum or number changing interactions. For particles which decouple from the rest of the Universe when non-relativistic (e.g. cold dark matter and baryons), the initial spectrum will be given by a Maxwell-Boltzmann distribution with occupation number

$$n^{\mathrm{MB}} \propto e^{-\frac{p^2}{2mT}}, \tag{1.66}$$

where m is the mass of the particle. Conservation of occupation number for non-relativistic particles with initial equilbrium spectrum given by the Maxwell-Boltzmann distribution implies that for these particles $T \propto p^2 \propto a^{-2}$.

The wavelength of a photon λ evolves as

$$\lambda \propto \frac{1}{p} \propto a \equiv \frac{1}{1+z}, \tag{1.67}$$

where we have defined the redshift z, the amount by which the wavelength of a visible photon shifts towards the red (as opposed to blue) part of the spectrum with the expansion. The terminology is derived from looking at visible light photons but of course the formula applies to all photons as well as other particles.

1.4.2 Friedmann equations

The Einstein's equations of general relativity, relating the metric tensor or the geometry of the Universe ($g_{\alpha\beta}$) to the stress energy tensor $T_{\mu\nu}$, are

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.68}$$

where the Einstein tensor is given by

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu}, \tag{1.69}$$

where $R_{\mu\nu}$ is the Ricci tensor

$$R_{\alpha\beta} = -\frac{\partial \Gamma^{\lambda}_{\lambda\alpha}}{\partial x^{\beta}} + \frac{\partial \Gamma^{\lambda}_{\alpha\beta}}{\partial x^{\lambda}} - \Gamma^{\lambda}_{\alpha\sigma}\Gamma^{\sigma}_{\beta\lambda} + \Gamma^{\lambda}_{\alpha\beta}\Gamma^{\sigma}_{\lambda\sigma} \tag{1.70}$$

and $R = R_{\alpha\beta}g^{\alpha\beta}$ is the Ricci scalar. The time-space components of the Ricci tensor, R_{0i} and R_{i0} are 3-vectors, a non-zero value for them would imply a preferred direction in space-time and must therefore vanish for an isotropic Universe.

The Christoffel symbols are easily calculated from Eq. 1.53 and the Friedmann metric, Eq. 1.48,

$$\begin{aligned}
\Gamma^{i}_{00} = \Gamma^{0}_{00} = \Gamma^{0}_{i0} &= \Gamma^{0}_{0i} = 0 \\
\Gamma^{0}_{ij} &= \frac{1}{2}g^{00}\left(-\frac{\partial g_{ij}}{\partial t}\right) = \frac{\dot{a}}{a}g_{ij} \\
\Gamma^{i}_{0j} &= \frac{1}{2}g^{ik}\frac{\partial g_{kj}}{\partial t} = \frac{\dot{a}}{a}\delta^{i}_{j}
\end{aligned} \tag{1.71}$$

We therefore have for the time-time component of the Ricci tensor,

$$R_{00} = -\frac{\partial \Gamma^{i}_{i0}}{\partial t} - \Gamma^{i}_{0j}\Gamma^{j}_{0i} = -3\left[\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^{2}\right] - \left(\frac{\dot{a}}{a}\right)^{2}\delta^{i}_{j}\delta^{j}_{i} = -3\frac{\ddot{a}}{a} \tag{1.72}$$

Calculation of the space-space components, R_{ij} is more involved. It can be simplified by using Cartesian (x^i) instead of spherical polar coordinates for the spatial part of the metric and doing the computation at the origin of the coordinate system [e.g. 25]. There is no loss of generality here, since we are in homogeneous and isotropic space and we can choose any point to be the origin of our coordinates. We can write the metric in Cartesian coordinates as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{1.73}$$

where we have defined the spatial metric (see Eq. 1.39)

$$\gamma_{ij} = \delta_{ij} + \frac{Kx_ix_j}{1-Kx^2} \tag{1.74}$$

At the origin of the coordinate system, $x^i = 0$, we have $\gamma_{ij} = \delta_{ij}$ and we can raise and lower the spatial indices at the origin using the Kronecker delta functions δ_{ij} and δ^{ij}. We will need the derivatives of the metric,

$$\frac{\partial \gamma_{ij}}{\partial x^k} = \frac{K}{1-Kx^2}\left(\delta_{ik}x_j + \delta_{jk}x_i\right) + \frac{2Kx_kx_ix_j}{(1-Kx^2)^2} \tag{1.75}$$

We note that the second term will vanish when we take another derivative to calculate R_{ij} at $x=0$ since it is second order in x_i. We can therefore ignore it. Similarly, we can also neglect the denominator in Eq. 1.75, giving

$$\frac{\partial\gamma_{ij}}{\partial x^k} = K\left(\delta_{ik}x_j + \delta_{jk}x_i\right) \tag{1.76}$$

We therefore have, using $g^{ij} = (1/a^2)\gamma^{ij}$ and $\gamma^{ij} \overset{x\to 0}{\approx} \delta^{ij}$,

$$\begin{aligned}
\Gamma^i_{jk} &= \frac{1}{2}g^{i\ell}\left[\frac{\partial g_{\ell j}}{\partial x^k} + \frac{\partial g_{\ell k}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^\ell}\right] \\
&= \frac{1}{2}\delta^{i\ell}\left[\frac{\partial \gamma_{\ell j}}{\partial x^k} + \frac{\partial \gamma_{\ell k}}{\partial x^j} - \frac{\partial \gamma_{jk}}{\partial x^\ell}\right] \\
&= \frac{K}{2}\delta^{i\ell}\left[\delta_{\ell k}x_j + \delta_{jk}x_\ell + \delta_{\ell j}x_k + \delta_{kj}x_\ell - \delta_{j\ell}x_k - \delta_{k\ell}x_j\right] \\
&= Kx^i\delta_{jk} \qquad (1.77)\\
\frac{\partial\Gamma^i_{jk}}{\partial x^\ell} &= K\delta^i_\ell\delta_{jk}. \qquad (1.78)
\end{aligned}$$

We also note that $\Gamma^i_{jk} \overset{x\to 0}{=} 0$. We therefore get for the space-space components of Ricci tensor at $x=0$,

$$\begin{aligned}
R_{ij} &\overset{x=0}{=} -\frac{\partial\Gamma^k_{ki}}{\partial x^j} + \frac{\partial\Gamma^0_{ij}}{\partial t} + \frac{\partial\Gamma^k_{ij}}{\partial x^k} - \Gamma^0_{ik}\Gamma^k_{j0} - \Gamma^k_{i0}\Gamma^0_{jk} + \Gamma^0_{ij}\Gamma^k_{0k} \\
&= -K\delta^k_j\delta_{ki} + \frac{\partial\dot{a}a}{\partial t}\delta_{ij} + K\delta^k_k\delta_{ij} - \dot{a}a\delta_{ik}\frac{\dot{a}}{a}\delta^j_k - \frac{\dot{a}}{a}\delta^k_i\dot{a}a\delta_{jk} + \dot{a}a\delta_{ij}\frac{\dot{a}}{a}\delta^k_k \\
&= \left[\frac{2K}{a^2} + \frac{\ddot{a}}{a} + \frac{2\dot{a}^2}{a^2}\right]\delta_{ij} \qquad (1.79)\\
&= \left[\frac{2K}{a^2} + \frac{\ddot{a}}{a} + \frac{2\dot{a}^2}{a^2}\right]g_{ij}(x=0), \qquad (1.80)
\end{aligned}$$

where in the last line we have repaced δ_{ij} with the metric g_{ij}, since they agree at $x=0$, lets call this point $\mathcal{P}$. Since this is a tensor equation, it is valid in any coordinate system. In particular, because of homogeneity and isotropy of space, we could have chosen a different point in space as the origin of our coordinate system and would have got the same result at the new origin. Moreover, the two coordinate systems are related by a cooridinate transoformation and in the new coordinate system the point $\mathcal{P}$ is at $x \neq 0$, but the tensor relation should remain valid at $\mathcal{P}$. Thus, the tensor equation derived above is valid for all x. The Ricci scalar is

$$
\begin{aligned}
R &= g^{ij}R_{ij} + g^{00}R_{00} \\
&= 6\left[\frac{K}{a^2} + \frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\right]
\end{aligned} \tag{1.81}
$$

We can now calculate the Einstein tensor,

$$
\begin{aligned}
G_{00} &= R_{00} - \frac{1}{2}Rg_{00} \\
&= 3\left[\frac{K}{a^2} + \frac{\dot{a}^2}{a^2}\right]
\end{aligned} \tag{1.82}
$$

$$
G_{ij} = -\left(\frac{K}{a^2} + \frac{2\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\right)g_{ij} \tag{1.83}
$$

Equating this with $8\pi G T_{\mu\nu}$, where the stress energy tensor is given by, for an ideal fluid, $T_{\mu\nu} = \rho U_\mu U_\nu + P\left(g_{\mu\nu} + U_\mu U_\nu\right)$ and evaluated in the local comoving coordinates. $U^\alpha = (1, \vec{0}), U_\alpha = (-1, \vec{0})$, we have $T_{00} = \rho$, $T_{ij} = Pg_{ij}$, where ρ is the total energy density at each point in the Universe as measured by comoving observers and P is the pressure.

The time-time component of the Einstein's equations give the Friedmann equation,

$$
H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{K}{a^2}. \tag{1.84}
$$

The spatial components give,

$$
\frac{2\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = -8\pi GP - \frac{K}{a^2} \tag{1.85}
$$

Combining it with the Friedmann equation, we obtain the Friedmann's second equation

$$
\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3P\right) \tag{1.86}
$$

The conservation of energy-momentum is built into the Einstein's equation. We can see this explicitly as follows. We can take the time derivative of Eq. 1.84 and combine it with Eq. 1.86 to get the mass conservation or continuity equation,

$$
\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + P\right) \tag{1.87}
$$

$$
\frac{\mathrm{d}\rho}{\mathrm{d}a} = -\frac{3}{a}\left(\rho + P\right) \tag{1.88}
$$

For pressure less matter ($P \ll \rho$) we have $\rho \propto a^{-3}$. For radiation, $P = \rho/3 \Rightarrow \rho \propto a^{-4}$. For cosmological constant, $P = -\rho \Rightarrow \rho = \text{constant}$.

The expansion rate today, at proper time t_0, is usually denoted by H_0, i.e. $H_0 \equiv H(t_0)$. The Hubble constant is usually measured in units of km $\text{s}^{-1}\text{Mpc}^{-1}$. It is also customary to define a dimensionless Hubble constant $h \equiv H_0/(100 \text{ km s}^{-1}\text{Mpc}^{-1})$. The units are chosen to reflect the order of magnitude of the typical distances and velocities of nearby galaxies. Since nearby objects follow Hubble law, $v = Hr$, where v is the radial velocity of a galaxy and r its distance from us, $h = 1$ implies that a galaxy 1 Mpc away is moving away from us with a speed of 100 km/s.

It is useful to define a critical energy-density, ρ_{cr}, which is the energy density of a flat Universe corresponding to the observed value of H_0. If $K = 0$, we have from Eq. 1.84

$$\rho_{\text{cr}} \equiv \frac{3H_0^2}{8\pi G}. \tag{1.89}$$

We can also divide the total energy density ρ into different kinds of fluids,

$$\rho = \rho_{\text{m}} + \rho_{\text{r}} + \rho_{\Lambda}, \tag{1.90}$$

where $\rho_{\text{m}} \propto a^{-3}$ is the energy density in non-relativistic matter, $\rho_{\text{r}} \propto a^{-4}$ is the energy density in radiation, and $\rho_{\Lambda} \propto a^0$ is the energy density in dark energy or cosmological constant Λ. In general we will use the notation ρ_i for energy density in species i. We will, for example, separate the energy density in baryons ρ_{b} and cold dark matter ρ_{cdm} which constitute matter, $\rho_{\text{m}} = \rho_{\text{b}} + \rho_{\text{cdm}}$ and also divide the radiation energy density into energy density in neutrinos, ρ_ν, and photons, ρ_γ. We define the relative density w.r.t critical density, Ω_i for species i as

$$\Omega_i = \frac{\rho_i(t_0)}{\rho_{\text{cr}}}, \tag{1.91}$$

where $\rho_i(t_0)$ is the energy density in species i today. With these definitions, the matter density at any time t corresponding to scale factor a can be written as $\rho_{\text{m}}(t) = \rho_{\text{m}}(t_0)a(t)^{-3}/a_0^{-3}$, where $a_0 \equiv a(t_0)$, and similarly for radiation and dark energy (and any other new species). With these definitions, we can write the Friedmann equation, Eq. 1.84, as

$$H^2 = H_0^2\left[\Omega_{\text{m}}\left(\frac{a_0}{a}\right)^3 + \Omega_{\text{r}}\left(\frac{a_0}{a}\right)^4 + \Omega_\Lambda + \Omega_{\text{K}}\left(\frac{a_0}{a}\right)^2\right], \tag{1.92}$$

where we have defined a Ω parameter for curvature term also,

$$\Omega_{\text{K}} \equiv \frac{-K}{a_0^2H_0^2}. \tag{1.93}$$

By definition, today at $t = t_0$, $H = H_0$ and $a = a_0$. Therefore, we see from Eq. 1.92 that

$$\Omega_{\rm m} + \Omega_{\rm r} + \Omega_\Lambda + \Omega_{\rm K} = 1, \tag{1.94}$$

and in particular for flat Universe, $\Omega_{\rm m} + \Omega_{\rm r} + \Omega_\Lambda = 1$.

We can solve the Friedmann equation analytically for simple cases when one of the fluid components dominates over all the others.

- Radiation domination

$$\begin{aligned}\frac{\dot{a}}{a} &= H_0 {\Omega_{\rm r}}^{1/2}\left(\frac{a_0}{a}\right)^2\\ a(t) &= a_0\left(2H_0{\Omega_{\rm r}}^{1/2}t\right)^{1/2}\end{aligned} \tag{1.95}$$

- Matter domination

$$\begin{aligned}\frac{\dot{a}}{a} &= H_0 {\Omega_{\rm m}}^{1/2}\left(\frac{a_0}{a}\right)^{3/2}\\ a(t) &= a_0\left(\frac{3}{2}H_0{\Omega_{\rm m}}^{1/2}t\right)^{2/3}\end{aligned} \tag{1.96}$$

- Dark energy domination

$$\begin{aligned}\frac{\dot{a}}{a} &= H_0 {\Omega_\Lambda}^{1/2}\\ a(t) &= a_0 e^{H_0{\Omega_\Lambda}^{1/2}(t-t_0)}\end{aligned} \tag{1.97}$$

- Curvature domination
  In a positive curvature Universe, $K = +1, \Omega_{\rm K} < 0$, we see from Eq. 1.94 that we must have one of the other components $\gtrsim \Omega_{\rm K}$ so that the sum is +1, i.e. $\Omega_{\rm m} + \Omega_{\rm r} + \Omega_\Lambda = 1 - \Omega_{\rm K} = 1 + |\Omega_{\rm K}|$ for $\Omega_{\rm K} < 0$. We therefore cannot have a positive curvature Universe that is dominated by $\Omega_{\rm K}$. However, a negatively curved Universe, $K = -1, \Omega_{\rm K} > 0$, in the absence of cosmological constant will eventually become curvature dominated, $\Omega_{\rm K} \to 1$ with solution $a(t) \propto t$.

Exercise 11

Solve the Friedmann equation for matter only Universe allowing for curvature (i.e. $\Omega_{\rm m} \neq 1$).

$$H^2 = H_0^2\left[\Omega_{\rm m}\left(\frac{a_0}{a}\right)^3 + \Omega_{\rm K}\left(\frac{a_0}{a}\right)^2\right]$$

You should get a solution similar to the one we got for the Newtonian cosmological model.
Hint: Do change of variables to conformal time (η), $\mathrm{d}t = a\mathrm{d}\eta$ and find the parameteric solutions $a(\eta)$ and $t(\eta)$.

Exercise 12

Einstein's blunder - Einstein tried to create a static Universe that would satisfy general relativity or equivalently Friedmann's equation. Friedmann equation in the presence of curvature is

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{K}{a^2} \tag{1.98}$$

Assume that only matter and cosmological constant are present. For $K = +1$, we can make the right hand side vanish at a single point in time (say at $a = a_0$) even without a cosmological constant by just fine tuning the energy density of matter ρ creating a Universe momentarily neither expanding or contracting, $\dot{a} = 0$. However, this is not enough to create a static Universe. We must also have vanishing acceleration. For a matter only Universe $\ddot{a} < 0$ (since $P = 0$) from Friedmann's second equation. Thus even if we can bring the Universe to rest by fine tuning matter density, it will start contracting in the future.
a) Show that by including cosmological constant and choosing ρ_{m} and ρ_Λ appropriately, acceleration can also be made to vanish in addition to the Hubble parameter.
b) Lets say we have done the above fine tuning and at $a = a_0$ both the Hubble and acceleration vanish. Show that this solution is unstable and any small perturbation will cause the Universe to expand or contract. Hint: Give a small perturbation to the Universe, e.g. $a \to a_0 + \delta a$ keeping the other variables same. Does the subsequent evolution (acceleration) bring the Universe back towards the equilibrium $a = a_0$ or takes it further and further away from it ?

1.5 Hubble law, distances, horizons

1.5.1 Proper distance and Hubble law

We are free to choose the origin of coordinates to be at any point in the homogeneous and isotropic Universe. To measure the *proper* or physical distance between the two points in the Friedmann Universe, lets choose as origin one of the points. The distance of the second point with coordinate $r = R$ and at constant time t is given by integrating the spatial part of the metric (Eq. 1.50)

$$\begin{aligned} d &= \int \left(g_{ij}\mathrm{d}x^i\mathrm{d}x^j\right)^{1/2} = \int_0^R g_{rr}\mathrm{d}r \\ &= a(t)\begin{cases} \sin^{-1} r, & K=+1, r<1 \\ \sinh^{-1} r, & K=-1 \\ r, & K=0 \end{cases} \\ &\equiv a(t)f(R) \end{aligned} \tag{1.99}$$

The *proper velocity* of the point at $r = R$ as seen by the observer at $r = 0$ is

$$\dot{d} = \dot{a}f(R) = \frac{\dot{a}}{a}d = Hd. \tag{1.100}$$

This is the Hubble law.

Interpreting the above equation as velocity = $H\times$ distance only makes sense for close objects, so that the scale factor does not change appreciably in the time it takes the light to travel from $r = 0$ to $r = R$, since we are evaluating H and proper distance d at a particular time t.[1]

We measure distances by observing light (or other causal) signals emitted from objects at $r = R$ at $r = 0$. The metric should not change in the time it takes to get from $r = R$ to $r = 0$, so that both points belong to the same Minkowski reference frame, i.e. the same frame should be locally flat to a good approximation at both points. Note that we get the same relation for all values of curvature, K. This makes sense, since in a small region of spacetime around $r = 0$ the flat spacetime is a good approximation and our measurements should not be able to see curvature of space. As we go further and further departures from the Hubble law become apparent. The proper distance is, in general, a calculable quantity but is not directly measurable over large distances.

Over cosmological distances, the definition of distance will change depending on what physical observable we use to measure the distance. Two useful ways to measure the distances on cosmological scales are: *standard candles* and *standard rulers* giving us the *luminosity distance* and the *angular diameter distance* respectively.

1.5.2 Luminosity distance

If we know the luminosity of an isotropic source, L (erg/s), integrated over some broad enough frequency band, the flux (erg/s/cm^2) measured at distance d is given

[1] We will see later that H^{-1} is approximately the size of the horizon.

by

$$F = \frac{L}{4\pi d^2}, \tag{1.101}$$

since the energy emitted by the source get distributed over an area of $4\pi d^2$ by the time it is measured by the observer. If we have a source of known luminosity and we measure its flux we get the luminosity distance d_L as

$$d_L = \left(\frac{L}{4\pi F}\right)^{1/2} \tag{1.102}$$

In an expanding Universe, the surface area of a sphere at coordinate distance r and at scale factor a_0 is given by integrating the angular part of metric (Eq. 1.48), $\int a_0^2 r^2 \mathrm{d}\Omega$, over all angles and is equal to $4\pi a_0^2 r^2$. Note that in the spherical polar coordinate system, the curvature enters only the radial part of the metric g_{rr} and not the angular part. The luminosity of a source at coordinate distance r (with us the observers at the origin of the coordinate system) will be divided over an area of $4\pi a_0^2 r^2$, where a_0 is the scale factor today, as this is the physical size of the wavefront today. In addition the energy of each photon is redshifted by a factor of $a_0/a_e = 1 + z$, where a_e is the scale factor when the photons are emitted. Thus the total energy received in the detector will be reduced by this factor. We should also take into account *cosmological time dilation*. Suppose two photons are emitted Δt_e apart in time at $a = a_e, t = t_e$ and we observe them today Δt_0 apart in time at $t = t_0, a = a_0$. Since the photons travel on null geodesics,

$$\mathrm{d}s^2 = 0. \tag{1.103}$$

There is no bending of light in homogeneous Universe, and the motion is purely radial with $\mathrm{d}\theta = \mathrm{d}\phi = 0$. The comoving distance traveled by the first photon is

$$\eta_1 = \int_{t_e}^{t_0} \frac{\mathrm{d}t}{a} = \int_0^r \frac{\mathrm{d}r}{1 - Kr^2} \tag{1.104}$$

Similarly for the second photon,

$$\begin{aligned}\eta_2 = \int_{t_e+\Delta t_e}^{t_0+\Delta t_0} \frac{\mathrm{d}t}{a} &= \int_0^r \frac{\mathrm{d}r}{1 - Kr^2} \\ &= \int_{t_e}^{t_0} \frac{\mathrm{d}t}{a} = \eta_1.\end{aligned} \tag{1.105}$$

We thus have,

$$
\begin{aligned}
0 &= \eta_2 - \eta_1 \\
&= \int_{t_e+\Delta t_e}^{t_0} \frac{\mathrm{d}t}{a} + \int_{t_0}^{t_0+\Delta t_0} \frac{\mathrm{d}t}{a} - \int_{t_e}^{t_0} \frac{\mathrm{d}t}{a} \\
&= -\int_{t_e}^{t_e+\Delta t_e} \frac{\mathrm{d}t}{a} + \int_{t_0}^{t_0+\Delta t_0} \frac{\mathrm{d}t}{a}
\end{aligned} \tag{1.106}
$$

Assuming Δt_e and Δt_0 are small so that the scale factor does not change between the emission of two photons, the integrals are readily evaluated and result is

$$
\begin{aligned}
\frac{\Delta t_e}{a_e} &= \frac{\Delta t_0}{a_0} \\
\frac{\Delta t_0}{\Delta t_e} &= \frac{a_0}{a_e} = 1 + z
\end{aligned} \tag{1.107}
$$

Thus the energy detected *per unit time* is smaller by additional factor of $1 + z$ coming from time dilation. Thus the detected flux is

$$
F = \frac{L}{4\pi a_0^2 r^2 (1+z)^2} \equiv \frac{L}{4\pi d_L^2}, \tag{1.108}
$$

where the last equality defines the luminosity distance in analogy with the formula for a non-expanding Universe,

$$
d_L = a_0 r(1+z). \tag{1.109}
$$

This can be compared with the comoving distance, $r_{\mathrm{cm}} = \eta$, which is defined as

$$
\eta = \int_0^r \frac{\mathrm{d}r'}{1 - Kr'^2} = \begin{cases} \sin^{-1} r, & K = +1, r < 1 \\ \sinh^{-1} r, & K = -1 \\ r, & K = 0 \end{cases} . \tag{1.110}
$$

In particular, the quantity that enters the luminosity distance is the coordinate distance and not the comoving distance.

1.5.3 Angular diameter distance

If we observe an object of size $R \ll d$ a distance d away, we have the small angle formula for the angular size (θ) of the object,

$$
\begin{aligned}
\theta &= \frac{R}{d} \\
d &= \frac{R}{\theta}.
\end{aligned} \tag{1.111}
$$

In Friedmann Universe, $g_{\theta\theta}$ tells us what is the *proper* or physical size of an object subtending an angle $\Delta\theta = \theta_R$. Note that we can always rotate the coordinate system so that $\Delta\phi = 0$, and the object is oriented θ direction.

$$R = \sqrt{g_{\theta\theta}}\Delta\theta = a_e r \theta_R, \tag{1.112}$$

where a_e is the scale factor when the photon was emitted as before. We define the angular diameter distance by comparing with Eq. 1.111 as

$$d_A = a_e r = \frac{a_0 r}{1+z} \tag{1.113}$$

The angular diameter distance and luminosity distance are related by

$$\frac{d_A}{d_L} = \frac{a_e r}{a_0 r(1+z)} = \frac{1}{(1+z)^2} \tag{1.114}$$

We note that for a flat Universe we can set $a_0 = 1$ and $a_e = 1/(1+z)$ but in general, the distances depend on the value of scale factors (a_e, a_0) and not just their ratio. However, the ratio d_L/d_A does not have this dependence on absolute values of a_0 or a_e, but only on their ratio.

The formulae for d_L and d_A depend on the coordinate system chosen, since the coordinate r appears explicitly. Lets instead choose the non-singular coordinate system defined by Eq. 1.45

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2\left[\mathrm{d}\Psi^2 + f(\Psi)^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right],$$
$$f(\Psi)^2 = \begin{cases} \sin^2\Psi, & K=+1, r<1 \\ \Psi^2, & K=0 \\ \sinh^2\Psi, & K=-1 \end{cases}. \tag{1.115}$$

The area of sphere at coordinate distance Ψ at scale factor a_0 is $4\pi f(\Psi)^2 a_0^2$ and the luminosity distance becomes

$$d_L = a_0 f(\Psi)(1+z). \tag{1.116}$$

The proper size of the object with angular size θ_R in this coordinate system is $R = \sqrt{g_{\theta\theta}}\theta_R = a_e f(\Psi)\theta_R$ and the angular diameter distance is

$$d_A = \frac{a_0 f(\Psi)}{1+z}. \tag{1.117}$$

The ratio of angular diameter distance to the luminosity distance remains $d_L/d_A = (1+z)^2$ and is independent of the coordinate system.

1.5.4 Comoving horizon or particle horizon

The fastest a particle can travel is with the speed of light i.e. on null geodesics with $\mathrm{d}s = 0$. Starting from *big bang*, $t = 0$, to any time t the distance (r_H) traveled by a massless particle is called the particle horizon or comoving horizon and decides the limits of observable Universe.

Looking at radial geodesics, without any loss of generality,

$$\begin{aligned} r_H(t) &= \int_0^r \frac{\mathrm{d}r}{1 - Kr^2} = \int_0^t \frac{\mathrm{d}t'}{a} = \eta \\ &= \int_0^a \frac{\mathrm{d}a'}{\dot{a}'a'} = \int_0^a \frac{\mathrm{d}a'}{H(a')a'^2} \end{aligned} \tag{1.118}$$

We can change variables from scale factor a to redshift z,

$$\begin{aligned} 1 + z &= \frac{a_0}{a} \\ \mathrm{d}z &= \frac{a_0}{a^2}\mathrm{d}a. \end{aligned} \tag{1.119}$$

We thus get,

$$r_H = a_0 \int_z^\infty \frac{\mathrm{d}z'}{H(z')} \tag{1.120}$$

Another useful form is,

$$r_H = \eta = \int_0^a \frac{\mathrm{d}\ln a'}{H(a')a'}, \tag{1.121}$$

where $1/(aH)$ is the comoving Hubble radius. It is the distance a particle travels in a logarithmic interval in scale factor. Since most of the contribution to the integral will come from the last $\ln a$ in a decelerating Universe, it is approximately the comoving horizon and $\ln a$ is approximately the scale over which H changes appreciably.

During radiation domination,

$$H \propto a^{-2}, aH \propto a^{-1}, \frac{1}{aH} \propto a, \tag{1.122}$$

during matter domination

$$H \propto a^{-3/2}, aH \propto a^{-1/2}, \frac{1}{aH} \propto a^{1/2}, \tag{1.123}$$

and during dark energy domination,

$$H \propto a^0, aH \propto a, \frac{1}{aH} \propto a^{-1}, \tag{1.124}$$

The physical horizon at any redshift is given by $a\eta \approx a/(aH) = 1/H$.

1.5.5 Event horizon

The (comoving) event horizon, $r_{\rm EH}$ is defined as the furthest (comoving) distance a particle starting today can travel in the future.

$$r_{\rm EH} = \int_{t_0}^{\infty} \frac{{\rm d}t'}{a(t')}. \tag{1.125}$$

In the absence of dark energy, the Universe today and in future would be matter dominated, $a \propto t^{2/3} \Rightarrow r_{\rm EH} \rightarrow \infty$. However, with cosmological constant dominating, we have $a = a_0 e^{H(t-t_0)}$ giving $r_{\rm EH} = 1/(a_0 H)$, i.e. the comoving distance a particle can travel even in infinite time is finite and is given by the horizon size today. For $t_2 > t_1$, we have

$$r_{\rm EH}(t_2) = \frac{1}{a(t_2)H} < \frac{1}{a(t_1)H} = r_{\rm EH}(t_1), \tag{1.126}$$

i.e. the event or future horizon decreases as the time passes in a Λ dominated Universe.

Chapter 2

Thermal history of the Universe

2.1 Equilibrium and freezeout

Lets take a particle with number density n. If there are no interactions, i.e. no processes which create or destroy particles, the particle number N in a volume V, $N = n.V \propto na^3$ is conserved. Therefore,

$$\begin{aligned} \frac{\mathrm{d}N}{\mathrm{d}t} = 0 = \frac{\mathrm{d}na^3}{\mathrm{d}t} &= a^3\frac{\mathrm{d}n}{\mathrm{d}t} + 3na^2\dot{a} \\ \Rightarrow \frac{\mathrm{d}n}{\mathrm{d}t} + 3Hn = 0. \end{aligned} \tag{2.1}$$

This is the continuity or number conservation equation in an expanding Universe. If we have processes creating or destroying particles, then we add these terms to the right hand side of Eq. 2.1.

$$\begin{aligned} \frac{\mathrm{d}n}{\mathrm{d}t} + 3Hn &= \text{ creation rate } - \text{ destruction rate} \\ &= \left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{\mathrm{cr}} - \left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{\mathrm{des}}. \end{aligned} \tag{2.2}$$

Equilibrium is achieved if the creation and destruction processes are much faster than the expansion rate, i.e.

$$\frac{1}{n}\left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{\mathrm{cr}} \sim \frac{1}{n}\left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{\mathrm{des}} \gg H. \tag{2.3}$$

In this case we can neglect the Hubble term in Eq. 2.2 compared to the creation and destruction terms and the instantaneous solution for the particle abundance is the same as it would be in a non-expanding Universe.

In the opposite case, when

$$H \gg \frac{1}{n}\left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{\mathrm{cr}} \sim \frac{1}{n}\left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{\mathrm{des}}, \tag{2.4}$$

Eq. 2.2 is well approximated by Eq. 2.1 and *freezeout* happens i.e. any creation and destruction processes are unimportant compared to the dilution of particle density due to expansion, and particle number is approximately conserved.

2.2 The Boltzmann equation and the Saha equation

The equilibrium phase space distribution, $f(p)$, of a particle depends on the statistics it follows and is given by

$$f(p) = \begin{cases} \frac{1}{e^{(E(p)-\mu)/T}-1} & \text{Bose} - \text{Einstein statistics} \\ \frac{1}{e^{(E(p)-\mu)/T}+1} & \text{Fermi} - \text{Dirac statistics} \\ e^{-(E(p)-\mu)/T} & \text{Boltzmann statistics} \end{cases}, \tag{2.5}$$

We will usually be interested in systems where particles have a large chemical potential, μ, or are non-relativistic, $E \sim m \gg T$. In this limit both the Fermi-Dirac and Einstein-Bose distributions reduce to Boltzmann distribution,

$$\frac{1}{e^{(E-\mu)/T} \pm 1} \overset{E-\mu \gg T}{\rightarrow} e^{\mu/T} e^{-E/T} \propto e^{-E/T}, \tag{2.6}$$

and we can assume that the paricles are described by Boltzmann statistics i.e. in this limit we can ignore the Pauli suppression/Bose enhancement factors (± 1) in the denominator as well as in the evolution equations. The chemical potential in this case determines the actual number density of particles.

Let us consider a 2-body process, with numbers labeling the particles,

$$1 + 2 \leftrightarrow 3 + 4. \tag{2.7}$$

The rate per unit volume for the forward process is $\langle \sigma v \rangle_{12} n_1 n_2$, where n_i is the number density of particle i, and $\langle \sigma v \rangle_{ij}$ is the velocity averaged cross section for particles i, j to intearct. The rate per unit for the backward process is $\langle \sigma v \rangle_{34} n_3 n_4$. For particle 1, we can write evolution equation as

$$\frac{\mathrm{d}n_1}{\mathrm{d}t} + 3Hn_1 = -\langle \sigma v \rangle_{12} n_1 n_2 + \langle \sigma v \rangle_{34} n_3 n_4, \tag{2.8}$$

and similarly for other particles if needed. Usually, we have one or more particles in equilibrium and described by their equilibrium distribution as a function of temperature and we are interested in one of the particles which goes out of equilibrium. Also, in general $\langle\sigma v\rangle_{12}$ is related to $\langle\sigma v\rangle_{34}$. In full thermal equilibrium the forward and backward rates must balance each other,

$$\langle\sigma v\rangle_{12} n_1^{\mathrm{eq}} n_2^{\mathrm{eq}} = \langle\sigma v\rangle_{34} n_3^{\mathrm{eq}} n_4^{\mathrm{eq}}, \tag{2.9}$$

where the equilibrium distributions n_i^{eq} are given in general by Fermi-Dirac or Bose-Einstein distributions depending on whether the particle is a fermion or a boson and by Boltzmann distribution in the approximation $E - \mu \gg T$. We can thus replace one of the cross sections in Eq. 2.8 using Eq. 2.9,

$$\frac{\mathrm{d}n_1}{\mathrm{d}t} + 3Hn_1 = \langle\sigma v\rangle_{12} n_1^{\mathrm{eq}} n_2^{\mathrm{eq}} \left[-\frac{n_1 n_2}{n_1^{\mathrm{eq}} n_2^{\mathrm{eq}}} + \frac{n_3 n_4}{n_3^{\mathrm{eq}} n_4^{\mathrm{eq}}} \right]. \tag{2.10}$$

This is one form of the Boltzmann equation and we can arrive at it more rigorously starting with the Boltzmann equation with the collison terms written as integral over the phase space densities.

As before, if $n_2\langle\sigma v\rangle_{12} \gg H$, the term in the brackets on right hand side must be vanishing giving

$$\frac{n_1 n_2}{n_1^{\mathrm{eq}} n_2^{\mathrm{eq}}} = \frac{n_3 n_4}{n_3^{\mathrm{eq}} n_4^{\mathrm{eq}}}. \tag{2.11}$$

This is the *Saha equation*. The Saha equation applies whenever the reaction rates are large. In particular, it is applicable even when some of the particles do not have the equilibrium Fermi-Dirac or Bose-Einstein distributions with zero chemical potential. Equilibrium abundances are functions of only temperature. For example, the electron abundance is decided by the baryon-anti-baryon asymmetry at temperatures $T \ll m_e$, where m_{e} is the electron mass and is thus independent of temperature once all positrons have annihilated. Electrons thus have non-equilibrium abundance or number density at temperatures below its mass. However, we can still have chemical equilibrium in reactions involving electrons and thus use the Saha equation e.g. initial stages of recombination.

For massive particles at temperature below their mass, $E = \sqrt{m^2 + p^2} \approx m + p^2/(2m) \gg T$, where p is the momentum and E is the energy, and for both fermions and bosons the abundance is given by integrating the phase space density over momentum (using natural units $\hbar = c = k_B = 1$),

$$n^{\rm eq} = g\int \frac{{\rm d}^3p}{(2\pi)^3}\frac{1}{e^{E/T}\pm 1} = g\int \frac{{\rm d}^3p}{(2\pi)^3}e^{-E/T} = g\left(\frac{mT}{2\pi}\right)^{3/2}e^{-m/T},\ \text{for}\, m \gg T.$$
$$= g\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{3/2}e^{-m/(k_BT)},\ \text{for}\, m \gg T, \tag{2.12}$$

where g is the degeneracy of the particle and in the last line we have restored physical constants (except speed of light which is still unity).

For massless particles, $E = p$, and we get

$$n^{\rm eq} = \begin{cases} g\frac{\zeta(3)T^3}{\pi^2} & \text{bosons},\ m \ll T \\ g\frac{3\zeta(3)T^3}{4\pi^2} & \text{fermions},\ m \ll T \\ g\frac{T^3}{\pi^2} & \text{Boltzmann statistics},\ m \ll T \end{cases}, \tag{2.13}$$

where ζ is the Riemann zeta function with $\zeta(3) = 1.20206$. Note that the last line in above equation gives the equilibrium number density for the massless particles following Boltzmann statistics. The fact that real particles are either bosons or fermions results in additional numerical factors of $\zeta(3)$ and $(3/4)\zeta(3)$ for bosons and fermions respectively.

2.2.1 Neutrino decoupling

For $\nu - e^-$ interactions, we have

$$\langle\sigma v\rangle = \sigma c = \sigma_{\rm wk} = (\hbar G_{\rm wk}k_BT)^2, \tag{2.14}$$

where $G_{\rm wk} = 1.16\times 10^{-5}$ GeV^{-2} is the weak interaction or Fermi coupling constant. The interaction rate, Γ_ν of neutrinos with electrons is therefore, with electron number density given by the equilibrium number density, $n^{\rm eq} \sim (k_BT/\hbar)^3$ (ignoring small numerical factors)

$$\Gamma_\nu = n_e\sigma_{\rm wk} \sim \frac{G_{\rm wk}^2(k_BT)^5}{\hbar} \tag{2.15}$$

The Hubble rate during radiation domination is given by

$$H \sim \sqrt{G\rho_r} \sim \left(\frac{G(k_BT)^4}{\hbar^3}\right)^{1/2}. \tag{2.16}$$

Comparing the weak interaction rate with the expansion rate we get,

$$\frac{\Gamma_\nu}{H} \approx G_{\rm wk}^2 \left(\frac{\hbar}{G}\right)^{1/2} (k_{\rm B}T)^3 \approx \left(\frac{T}{10^{10}\,{\rm K}}\right)^3 \approx \left(\frac{T}{1\,{\rm MeV}}\right)^3 \tag{2.17}$$

Thus, below 1 MeV, the weak interaction rate becomes slower compared to the expansion rate of the Universe and neutrinos decouple from the electron-positrons. However, this decoupling temperature is still larger compared to the electron mass, $m_{\rm e} = 0.511$ MeV. Thus the electron-positrons are still numerous and in equilibrium when the neutrinos decouple. The electron-positron annihilation happens much later after neutrino decoupling and therefore most of the energy from electron-positron annihilation goes into the photons.

Before neutrino decoupling, at $T \gtrsim 1$ MeV, we have the following relativistic degrees of freedom:

- Fermions: $2e^- + 2e^+ + 2 \times 3\nu$
- Bosons: 2γ

The entropy density, s, is given by the second law of thermodynamics,

$$\mathrm{d}S = \frac{\mathrm{d}E + P\mathrm{d}V}{T}, \tag{2.18}$$

where $S = sV$ is the total entropy in volume V, E is the total thermal energy and P is the pressure. We can rewrite the above equation in terms of entropy density and energy density ($\rho = E/V$),

$$\begin{aligned} \mathrm{d}(sV) &= \frac{\mathrm{d}(\rho V) + P\mathrm{d}V}{T} \\ TV\left.\frac{\partial s}{\partial T}\right|_V \mathrm{d}T + Ts\mathrm{d}V &= V\left.\frac{\partial \rho}{\partial T}\right|_V \mathrm{d}T + (\rho + P)\,\mathrm{d}V. \end{aligned} \tag{2.19}$$

Evaluating the above at constant temperature, we can equate the coefficients of dV giving,

$$s = \frac{\rho + P}{T}. \tag{2.20}$$

Alternatively, we have the Gibbs free energy given by

$$\phi = N\mu = E - TS + PV = 0 \tag{2.21}$$

if the chemical potential $\mu = 0$ giving again Eq. 2.20. We thus have an expression for entropy density at a fixed temperature.

We can also equate the coefficients of dT at constant volume, giving

$$T\frac{\partial s}{\partial T} = \frac{\partial \rho}{\partial T}. \tag{2.22}$$

Using Eq. 2.20 for s, we get a differential equation for P,

$$\frac{\partial P}{\partial T} = \frac{\rho + P}{T}. \tag{2.23}$$

We can solve this formally by multiplying by the integrating factor, $1/T$,

$$\frac{1}{T}\frac{\partial P}{\partial T} - \frac{P}{T^2} = \frac{\rho(T)}{T^2}$$

$$P = T\int \frac{\rho}{T^2}\mathrm{d}T \tag{2.24}$$

The energy density is given by integral of particle energy over the distribution function,

$$\rho = \int \frac{\mathrm{d}^3 p}{(2\pi)^3} f(p)\sqrt{p^2 + m^2} \tag{2.25}$$

For massless particles, we get for bosons at temperature T

$$\rho = \frac{1}{2} g_{\mathrm{B}} a_{\mathrm{R}} T^4 \tag{2.26}$$

where g_{B} is the total bosonic degrees of freedom and $a_{\mathrm{R}} = \frac{8\pi^5 {k_{\mathrm{B}}}^4}{15c^3h^3}$ is the radiation constant. For Fermions we have

$$\rho = \frac{7}{8}\frac{1}{2} g_{\mathrm{F}} a_{\mathrm{R}} T^4 \tag{2.27}$$

From Eq. 2.24 we get for both bosons and fermions,

$$P = \frac{1}{3}\rho, \tag{2.28}$$

and therefore for entropy density we have

$$s = \frac{\rho + P}{T} = \begin{cases} \frac{2}{3} g_{\mathrm{B}} a_{\mathrm{R}} T^3 : & \text{bosons} \\ \frac{7}{12} g_{\mathrm{F}} a_{\mathrm{R}} T^3 : & \text{fermions} \end{cases}, \tag{2.29}$$

Since $e^- e^+$ annihilation happens in equilibrium, except for the very last stages, total entropy is conserved for the electromagnetic plasms. Denoting the quantities before annihilation by primes ($'$), we have the total entropy before annihilation at scale factor a' and temperature T' equal to the total entropy after annihilation in the same comoving volume at scale factor a and temperature T,

$$s'(T')a'^3 = s(T)a^3 \tag{2.30}$$

Before annihilation, we have contribution from e^-e^+ ($g_{\rm F} = 4$) and photons ($g_{\rm B} = 2$) giving

$$\begin{aligned} s'(T')a'^3 &= \left(4.\frac{7}{12}a_{\rm R}T'^3 + 2\frac{2}{3}a_{\rm R}T'^3\right)a'^3 \\ &= \frac{11}{3}a_{\rm R}T'^3a'^3. \end{aligned} \tag{2.31}$$

After annihilation all entropy is in photons,

$$s(T)^3a^3 = \frac{4}{3}a_{\rm R}T^3a^3. \tag{2.32}$$

Using Eq. 2.30 we get

$$\left(\frac{T}{T'}\right)^3 = \frac{11}{4}\left(\frac{a'}{a}\right)^3 = \left(\frac{T}{T'_\nu}\right)^3, \tag{2.33}$$

where the last equality follows since the neutrino temperature T'_ν is equal to the temperature of the electromagnetic plasma (T') before electron-positron annihilation when all particle species are in equilibrium with each other. Since neutrinos are decoupled during the e^-e^+ annihilation, their entropy is separately conserved with neutrino temperature just redshifting $T_\nu \propto 1/a$. The neutrino temperature before and after the annihilation is thus simply related by

$$\frac{T_\nu}{T'_\nu} = \frac{a'}{a}. \tag{2.34}$$

Using it in Eq. 2.33 to replace the scale factors, we get the following relation between the neutrino and photon temperatures after the electron-positron annihilation is complete:

$$\frac{T_\nu}{T} = \left(\frac{4}{11}\right)^{1/3}. \tag{2.35}$$

In standard cosmology, there is no significant event that can change the temperature of neutrinos or photons after the epoch of electron positron annihilation and this relationships persists until today. The CMB temperature today is measured to be

$T_{\rm CMB} = 2.725$ K giving the massless neutrino (cosmic neutrino background or CNB) temperature today to be $T_{\rm CNB} = 1.945$ K $= 1.676 \times 10^{-4}$ eV.

This is however not thw whole story. Since e^-e^+ annihilation happens quite close to the neutrino decoupling, some of the energy does go into neutrinos, i.e. neutrino decoupling is not 100% at the time of e^-e^+ annihilation. Some of the e^-e^+ do annihilate into the neutrinos increasing their energy. Also as e^-e^+ annihilate, the plasma temperature is raised above that of neutrinos. The elastic scattering of neutrinos on hotter electrons and positrons also heats them up, transferring small amount of energy to the neutrinos from e^-e^+. The total energy density in neutrinos without these corrections is given by (Eq. 2.27)

$$\rho_\nu = N_\nu \frac{7}{8} a_{\rm R} T_\nu^4, \qquad (2.36)$$

where $N_\nu = 3$ is the number of flavors of standard neutrinos. We take into account the corrections outlines above by replacing N_ν with $N_\nu^{\rm eff}$, also known as the effective relativistic degrees of freedom in cosmology literature. A lot of work has gone into finding the precise value of $N_\nu^{\rm eff}$ taking into account the corrections outlined above as well as the effect of electromagnetic interactions in calculating the total energy density of electromagnetic plasma (compared to our claculation which treats electrons, positrons and photons as independent particles and just adds up theier entropy). These calculations give $N_\nu^{\rm eff} = 3.046$, or a $\sim 1\%$ level difference from our simple calculation [26]. Most cosmological observables are just sensitive to the total energy density of decoupled or free streaming relativistic particles e.g. CMB anisotropies. Usually if we have new relativistic particles, e.g. dark radiation, which are decoupled from the baryon-photon plasma, it is clubbed with neutrinos and their effect in cosmological calculations taken into account by adding the contribution of new relativistic particles to the radiation energy density to $N_\nu^{\rm eff}$. Since neutrinos participate in nuclear reactions with energy dependent cross sections, the exact spectrum of neutrinos is important for precise calculations of BBN. The extra energy added to neutrinos during e^-e^+ annihilation goes into high energy tail of the original Fermi-Dirac distribution distorting the spectrum of neutrinos from a thermal spectrum and BBN is sensitive to these distortions of the neutrino spectrum.

2.2.2 Big bang nucleosynthesis (BBN)

After the neutrons and protons condense out of the quark-gluon plasma at $T \lesssim 100\,\text{MeV}$, they are initially in equilibrium with nuclear reactions changing neutrons to protons and vice versa.

$$\begin{aligned} n + \nu_{\mathrm{e}} &\leftrightarrow p + e^- \\ p + \bar{\nu}_{\mathrm{e}} &\leftrightarrow n + e^+ \\ n &\rightarrow p + e^- + \bar{\nu}_{\mathrm{e}} \end{aligned} \tag{2.37}$$

The mass difference between the proton and neutron is $\Delta m = m_{\mathrm{n}} - m_{\mathrm{p}} = 1.29$ MeV

In equilibrium we can use the Saha equation. We can use Boltzmann statistics for protons and neutrons since protons and neutrons are non-relativistic and equilibrium number densities for neutrinos and electrons and the corresponding anti-particles, reducing the Saha equation to

$$\frac{n_{\mathrm{p}}}{n_{\mathrm{p}}^{\mathrm{eq}}} = \frac{n_{\mathrm{n}}}{n_{\mathrm{n}}^{\mathrm{eq}}}. \tag{2.38}$$

We can ignore the small mass difference between the mass of proton and neutrons in the prefactor multiplying the exponential in Eq. 2.12 where it makes negligible difference but keep it in the exponential, giving

$$\frac{n_{\mathrm{n}}}{n_{\mathrm{p}}} = \frac{g_{\mathrm{n}}}{g_{\mathrm{p}}} e^{-\Delta m/T} = e^{-1.29\ \mathrm{MeV}/T}. \tag{2.39}$$

At $T \gg \Delta m$ we have equal number densities for protons and neutrons $n_{\mathrm{n}} = n_{\mathrm{p}}$, while at $T \ll \Delta m$ the neutron number density will become negligible, $n_{\mathrm{n}}/n_{\mathrm{p}} \rightarrow 0$ if equilibrium was maintained. However, these are weak scale interactions which will freezeout at $T \approx 1$ MeV. At freezeout, $n_{\mathrm{n}}/n_{\mathrm{p}} \approx e^{-1.29} \approx 0.275$. A more precise calculation solving the actual evolution equation for abundance of neutrons, with $n = n_{\mathrm{n}}$ in Eq. 2.2 including the neutron decay in the destruction term, yields $n_{\mathrm{n}}/n_{\mathrm{p}} \approx 1/6$ or relative abundance of neutrons $X_{\mathrm{n}} \equiv n_{\mathrm{n}}/(n_{\mathrm{n}} + n_{\mathrm{p}}) \approx 1/7$. After the freezeout, the proton is stable but neutron decays with a lifetime of $\tau_{\mathrm{n}} \approx 885.7 \pm 0.8$ s. The neutron number density evolves due to decay (ignoring dilution ($\propto a^{-}$) due to the expansion of the Universe) according to

$$\frac{\mathrm{d}n_{\mathrm{n}}}{\mathrm{d}t} = \frac{-n_{\mathrm{n}}}{\tau_{\mathrm{n}}} \tag{2.40}$$

and the neutron to baryon number ratio (or relative abundance of neutrons) drops as

$$X_{\mathrm{n}} = X_{\mathrm{n}}^{i} e^{-\Delta t/\tau_{\mathrm{n}}}, \tag{2.41}$$

where Δt is the time after freezeout, and X_n^i is the initial neutron fraction at freezeout. Note that in the ratio of densities, the dilution due to expansion ($\propto a^{-3}$) drops out and Eq. 2.41 is valid in the expanding Universe as well. Fitting Eq. 2.41 to numerical results gives $X_n^i \approx 0.15 \approx 1/7$.

The first step in trying to fuse neutrons and protons to form nuclei heavier than hydrogen is formation of deuterium.

$$n + p \rightarrow d + \gamma,\ E_\gamma = 2.22\ \text{MeV}. \tag{2.42}$$

The reverse process photo-dissociates deuterium,

$$d + \gamma \rightarrow n + p \tag{2.43}$$

if the photon has energy greater than 2.22MeV.

Exercise 13

What is the ratio of photons with energy $E_\gamma \geq 2.22$ MeV to the baryon number density at $T = 0.07$ MeV. Hint: Take the value of baryon density parameter from Planck 2018 results of $\Omega_b h^2 = 0.0224$. Also calculate the total baryon to photon ratio, $\eta = n_B/n_\gamma$, after electron-positron annihilation, where n_B is the number density of baryons (protons + neutrons) and n_γ is the number density of photons.

Even at $T \ll 2.22$ MeV, there are many photons per baryon with energy $\gtrsim$ 2.22 MeV, which can destroy deuterium. This is because the photon to baryon ratio, $1/\eta \equiv n_\gamma/n_B \approx 1.47 \times 10^9 \ggg 1$. We must wait until deuterium destroying photons per baryon becomes much less than unity before we can start forming non-negligible amount of deuterium. This is also known as the deuterium bottleneck since until we form deuterium we cannot start to form heavier nuclei.

Therefore, nucleosynthesis can only start after $e^{-2.22\ \text{MeV}/T} \lesssim 10^{-9}$ or $T \lesssim 2.22/21$ MeV ≈ 0.1 MeV. The photo-dissociation of deuterium becomes negligible at $T \approx 2.22$ MeV$/30 \approx 0.07$ MeV and nucleosynthesis can start. This corresponds to proper time $t \approx 200$ s and at this time we have the neutron fraction of approximately $X_n \approx 1/7 e^{-200/885.7} \approx 0.11$.

Once the high energy photons disappear, the reaction 2.42 is very fast and consumes all neutrons. The nuclear reactions then proceed further creating heavier elements. The important nuclear reactions for the BBN are

$$\begin{aligned} d + p &\rightarrow {}^3\text{He} + \gamma \\ d + d &\rightarrow {}^3\text{H} + p \\ d + d &\rightarrow {}^3\text{He} + n \\ {}^3\text{H} + d &\rightarrow {}^4\text{He} + n \\ {}^3\text{He} + d &\rightarrow {}^4\text{He} + p \\ {}^3\text{H} + {}^4\text{He} &\rightarrow {}^7\text{Li} + \gamma \\ {}^3\text{He} + {}^4\text{He} &\rightarrow {}^7\text{Be} + \gamma \end{aligned} \tag{2.44}$$

The relative amounts of different elements produced in this chain of reactions depends on the relative binding energies, E_{bind} of different elements, which is the difference between the sum of individual energies of nucleons and the total energy of nucleus,

$$E_{\text{bind}} = N_{\text{p}} m_{\text{p}} + N_{\text{n}} m_{\text{n}} - m_{\text{A}} \text{ amu}, \tag{2.45}$$

where N_{p} and N_{n} are the number protons and neutrons respectively in the nucleus and m_{A} is the atomic mass number or the mass of the nucleus and amu = 931.494 MeV is the atomic mass unit.

Among the light elements, ^{4}He has the highest binding energy per nucleon, much higher than $d, {}^3\text{He}, {}^3\text{He}, {}^7\text{Li}$, and ^{7}Be. This has important implications for BBN abundances. Once we form ^{4}He, it is very hard to destroy and all nucleons want to end up in ^{4}He. This is in fact what happens, BBN produces maximum amount of ^{4}He possible putting all of the neutrons in ^{4}He. Since each ^{4}He nucleus has two neutrons, the number density of $^4\text{He} \approx 1/2X_{\text{n}}$, and since the mass of helium is 4× that of neutron or proton, the mass abundance of helium-4, $Y_{\text{He}} \equiv \rho_{4\text{He}}/\rho_{\text{b}} \approx 2X_{\text{n}} \approx 0.22$. This rough estimate is in fact quite accurate and a more precise calculation yields $Y_{\text{He}} \approx 0.24$. The remaning protons, which did not go into ^{4}He make up hydrogen. In addition BBN produces trace amounts of other elements, $d, {}^3\text{He}$, and ^{7}Li and ^{6}Li. The other two nuclei, ^{7}Be and ^{3}H are unstable and decay into ^{7}Li and ^{3}He respectively soon after the BBN is over.

Mass gap at atomic mass of 5 and 8

There is no stable nucleus with mass 5 or 8 and this means that BBN cannot build up heavy nuclei by adding a proton to helium or combining two helium nuclei. The coulomb barrier is too great to add protons or other nuclei to $^7\text{Li}, {}^7\text{Be}$. Thus BBN produces very negligible amount of elements heavier than atomic mass 7. This bottleneck is overcome in stars by 3-body reactions, ^{4}He + ^{4}He + ^{4}He $\rightarrow^{12}$ C to produce carbon. These reactions need high densities such as in the cores of stars in addition to an enhancement of the reaction rates by resonance. Density in the early Universe is too low for these reactions to be happen.

Open questions in BBN

The particle data group (PDG) reviews are an excellent source for the current status of the field of BBN. We saw that the abundance of light elements, d, ^{4}He, ^{3}He, ^{7}Li, ^{6}Li is governed by the baryon to photon ratio, η. This is the only cosmological parameter which enters the BBN calculations. The abundaces of course depends sensitively on the nuclear reaction rates which have some uncertainties. The helium-4 is the sink for all neutrons and helium abundance has very little sensitivity to the nuclear reaction rates.

We can measure the abundance of these elements today in the gas and stars in our own Galaxy as well as external galaxies. However, stars also create these elements in their cores through nuclear reactions (such as helium) as well as destroy elements. In particular both lithium and deuterium are destroyed in stars since these are fragile elements with low binding energy. This processing of primordial gas and modification of abundances has to be taken into account when measuring the abundance of these elements in the Universe today. It is particularly hard to measure the abundance of ^{3}He since it can be both created and destroyed in stars. The BBN prediction for the ratio of number density ^{3}He to hydrogen is $^3\mathrm{He}/H \approx 10^{-5}$. The deuterium abundance, D/H, agrees very well between the theory and observations with $\mathrm{D/H} \approx 2.6 \times 10^{-5}$.

The most significant discrepancy between theory and observations for BBN is in the observed and predicted abundance of lithium. Lithium is a fragile element which is only destroyed in stars and the source of all lithium in existence today is BBN. The observed value of ratio of number density of lithium to hydrogen is $(\mathrm{Li/H})^{\mathrm{obs}} = 1.6 \pm 0.3 \times 10^{-10}$ and theoretical prediction is $(\mathrm{Li/H})^{\mathrm{theory}} = 5.3 \times 10^{-10}$ [27–29][1]. This might be an indication of misinterpretaion of data and in particular something missing in our understanding of astrophysics of stars or it could be an indication of new physics beyond the standard model which modfies the theoretical prediction of primordial Li abundance. The astrophysical solution to the lithium problem invokes destruction of lithium in the cores of stars and transfer of this reprocessed material to the stellar surface through convection. Alternatively, new physics, such as injection of energetic photons in the early Universe from decay of dark matter, can destroy lithium produced in the BBN bringing its abundance

[1] See particle data group review for the current status `http://pdg.lbl.gov/`

down to the measured one. The challenge in these solution is to leave the other abundances unchanged, especially the deuterium abundance which agrees so well between the theory and observations.

2.3 Thermal production of dark matter

One of the ways to produce cold dark matter, a new massive particle beyond the standard model, is by thermal production. We note that the neutrinos, since they have a small mass, are non-relativistic today, and are decoupled from the baryons-photon fluid at $T \lesssim 1$ MeV, will also be classified as dark matter today. However, since they decoupled when they were relativistic, and became non-relativistic only recently, they are classified as hot dark matter. In particular, neutrinos cannot start clustering gravitationally while they are relativistic, moving around with the speed of light. Whereas, we need dark matter to start clustering much before the epoch of recombination, at $z \gg 1000$, in order to explain the large scale structure of the Universe today. We therefore need the dominant component of dark matter to be *cold*. The cold dark matter, by definition, must be non-relativistic at the time of their decoupling from the standard model particles, at $T \gtrsim$ MeV. Thus the dark matter must interact *weakly* with the standard model particles, either through the standard model weak force or a new force of similar strength. Such dark matter candidates are also called Weakly Interacting Massive particles or WIMPs. A comprehensive review of WIMP models of dark matter is given in [30].

Lets us suppose that in the early Universe the dark matter particles (χ) are in chemical equilibrium with the standard model particles (denoted by SM) through creation and destruction processes,

$$\chi + \bar{\chi} \leftrightarrow \mathrm{SM} + \mathrm{SM} \tag{2.46}$$

If there are equal number of dark matter and anti-dark matter particles (or if χ is its own anti-particle), then in equilibrium the chemical potential will vanish and they will have equilibrium Bose-Einstein or Fermi-Dirac distributions ($f(p)$). At temperature T much larger compared to the mass m of the dark matter particle the number density will $n_\chi^{\mathrm{eq}} \propto T^3$. At $T \ll m$, $f(p) \approx e^{-(m^2+p^2)^{1/2}/T}$ and from Eq. 2.12

$$n_\chi^{\mathrm{eq}} \approx g_\chi \left(\frac{mT}{2\pi\hbar^2}\right)^{3/2} e^{-m/T}. \tag{2.47}$$

If dark matter remained in equilibrium with the standard model particles, its number density will decrease exponentially until all DM has annihilated, much like the electrons-positrons. However, the dark matter annihilation process goes out of equilibrium and freezes-out because the number density of dark matter particles decreases due to the expansion of the Universe. Assuming that the dark matter is its own anti-particle ($\chi = \bar{\chi}$), the annihilation rate per unit volume is $\langle\sigma v\rangle_{\chi\chi} n_\chi^2/2$, where the factor of 2 corrects for double counting (i.e. particle 1 annihilating with particle 2 is same as particle 2 annihilating with particle 1). However, the creation and annihilation terms should be multiplied by 2 in the Boltzmann equation, since each creation and annihilation produces or destroys two particles and will cancel the factor of 1/2 from double counting. The evolution of dark matter number density is thus given by, using Eq. 2.10,

$$\frac{\mathrm{d}n_\chi}{\mathrm{d}t} + 3Hn_\chi = -\langle\sigma v\rangle_{\chi\chi}\left(n_\chi^2 - (n_\chi^{\mathrm{eq}})^2\right), \tag{2.48}$$

where we have used the fact that the other two SM particles are in full thermal equilibrium. Freezeout happens when annihilation rate $\Gamma_{\chi\chi} = n_\chi\langle\sigma v\rangle_{\chi\chi}$ equals the Hubble expansion rate, H. In the early radiation dominated Universe,

$$H = \left(\frac{8\pi G\rho}{3}\right)^{1/2} = \left(\frac{4\pi G g_* a_{\mathrm{R}} T^4}{3}\right)^{1/2}, \tag{2.49}$$

where

$$g_* = g_{\mathrm{B}} + \frac{7}{8} g_{\mathrm{F}} \tag{2.50}$$

is the total effective relativistic degrees of freedom, g_{B} is the total bosonic degrees of freedom, and g_{F} is the total fermionic degrees of freedom, and $\frac{1}{2}a_{\mathrm{R}}T^4$ is the energy density for one bosonic degree of freedom. The quantity g_* is a function of temperature and includes all particles whose mass is smaller compared to the temperature. As we go further back in time, the temperature increases and it becomes kinematically possible to produce more and more massive particles. For example, at $\gtrsim 100$ MeV it becomes possible to produce muons-anti muons and they must be counted in g_*. At even higher temperature we can produce pions and so on. Using definition of the reduced Planck mass,

$$M_{\mathrm{P}} = \left(\frac{1}{8\pi G}\right)^{1/2} = \left(\frac{\hbar c}{8\pi G}\right)^{1/2} = 2.435 \times 10^{18}\mathrm{GeV}, \tag{2.51}$$

also in natural units $a_{\mathrm{R}} = \pi^2/15$ giving

$$H = \frac{\pi}{3\sqrt{10}}\frac{g_*^{1/2}T^2}{M_{\mathrm{P}}} = 0.33\frac{g_*^{1/2}T^2}{M_{\mathrm{P}}} \tag{2.52}$$

For dark matter particles with mass m_χ, the energy density of dark matter at any time is $\rho_{\mathrm{cdm}} = n_\chi m_\chi$. Since after freezeout the dark matter number is conserved, the ratio of dark matter number density to total entropy density of the Universe is also conserved after freezeout since the standard model sector evolves in equilibrium. Thus,

$$\frac{n_\chi(T_\mathrm{f})}{s(T_\mathrm{f})} = \frac{n_\chi(T_0)}{s(T_0)} \tag{2.53}$$

where T_f is the freezeout temperature. At freezeout, $s(T_\mathrm{f}) = (2/3)g_* a_\mathrm{R} T^3$ and together with the condition $\Gamma_{\chi\chi} = H$ at freezeout we have

$$\begin{aligned}\frac{n_\chi(T_\mathrm{f})}{s(T_\mathrm{f})} &= \frac{H}{\langle\sigma v\rangle_{\chi\chi}} \frac{3}{2g_* a_\mathrm{R} {T_\mathrm{f}}^3} \\ &\approx \frac{0.75}{\langle\sigma v\rangle_{\chi\chi} g_*^{1/2} M_\mathrm{P} T_\mathrm{f}}\end{aligned} \tag{2.54}$$

Today, with CMB temperature at T_0, and neutrino temperature slightly smaller,

$$n_\chi(T_0) = \frac{\rho_\mathrm{cdm}}{m_\chi} = \frac{\Omega_\mathrm{cdm}\rho_\mathrm{cr}}{m_\chi} = \frac{3\Omega_\mathrm{cdm} H_0^2 {M_\mathrm{P}}^2}{m_\chi}. \tag{2.55}$$

The total entropy density today is the sum of entropy density in photons and neutrinos,

$$s(T_0) = \left(\frac{4}{3} + \frac{7}{2}\frac{4}{11}\right) a_\mathrm{R} T_0^3 \approx 1.71 T_0^3 = 2.2 \times 10^{-38}\ \mathrm{GeV}^3 \tag{2.56}$$

where $T_0 = 2.725$ K $= 2.35 \times 10^{-4}$ eV. Also, writing $H_0 = 100$ km/s/Mpc $h = 2.1 \times 10^{-42}\ h$ GeV, $H_0^2 {M_\mathrm{P}}^2 = 2.7 \times 10^{-47}\ h^2\ \mathrm{GeV}^4$. Putting it all together with Eq. 2.53, we get

$$\begin{aligned}\frac{0.75}{\langle\sigma v\rangle_{\chi\chi} g_*^{1/2} M_\mathrm{P} T_\mathrm{f}} &= \frac{\Omega_\mathrm{cdm} h^2 3.65 \times 10^{-9}\ \mathrm{GeV}}{m_\chi} \\ \Omega_\mathrm{cdm} h^2 &= \frac{8.45 \times 10^{-11}\ \mathrm{GeV}^{-2}}{\langle\sigma v\rangle_{\chi\chi}} \frac{m_\chi}{T_\mathrm{f}} \\ &= \frac{9.86 \times 10^{-28}\ \mathrm{cm}^3/\mathrm{s}}{\langle\sigma v\rangle_{\chi\chi} g_*^{1/2}} \frac{m_\chi}{T_\mathrm{f}}\end{aligned} \tag{2.57}$$

Since before the freezeout, the dark matter particles follow equilibrium distribution, We can find the freezeout temperature by using n_χ^eq in the freezeout condition

$\Gamma_{\chi\chi} = H$, giving the equation,

$$\langle\sigma v\rangle_{\chi\chi} g_\chi \left(\frac{m_\chi T_{\rm f}}{2\pi}\right)^{3/2} e^{-m_\chi/T_{\rm f}} = 0.33 \frac{g_*^{1/2} {T_{\rm f}}^2}{M_{\rm P}}$$
$$\left(\frac{m_\chi}{T_{\rm f}}\right)^{1/2} e^{-m_\chi/T_{\rm f}} = \frac{5.2 g_*^{1/2}}{g_\chi m_\chi M_{\rm P} \langle\sigma v\rangle_{\chi\chi}}$$
$$\frac{m_\chi}{T_{\rm f}} \approx 40.7 + \ln\left[\frac{g_\chi m_\chi \langle\sigma v\rangle_{\chi\chi}}{g_*^{1/2}} \left(\frac{m_\chi}{T_{\rm f}}\right)^{1/2}\right], \quad (2.58)$$

where all quantities inside the logarithm are in units of GeV. If we put typical weak scale cross section from Eq. 2.14, at $T \approx 10$ GeV, the thermally averaged cross section is $\sim 1.3 \times 10^{-8}$ GeV^{-2} = 10^{-25}cm^3/s, $g_*(T_{\rm f}) \approx 90$ which is approximately the degeneracy in the standard model for temperatures 10 GeV $\lesssim T_{\rm f} \lesssim$ 100 GeV and use $m_\chi \sim$ 300GeV, $g_\chi = 2$, we get $m_\chi/T_{\rm f} \approx 28$ consistent with our adopted values of m_χ and $T_{\rm f}$ and

$$\Omega_{\rm cdm} h^2 \approx \frac{2.9 \times 10^{-27}\ {\rm cm}^3/{\rm s}}{\langle\sigma v\rangle_{\chi\chi}}. \quad (2.59)$$

Therefore the cross section needed for getting $\Omega_{\rm cdm} h^2 \approx 0.1$ is $\langle\sigma v\rangle_{\chi\chi} \approx 3 \times 10^{-26}$ cm^3/s. Thus GeV mass particles with weak scale cross sections naturally produce the correct abundance of dark matter. Such particles are called WIMPs and this coincidence was termed the *WIMP miracle*.

If the dark matter is indeed a WIMP, this implies weak scale interactions with the standard model particles must exist even today. In particular, WIMP annihilation to standard model particles in our own Galaxy as well as external galaxies will result in gamma-ray emission (*indirect detection*). Elastic scattering of dark matter passing through earth with standard model particles like electrons and nuclei can be detected in the laboratory (*direct detection*). However, years of direct and indirect searches for the WIMPs have failed to detect. In particular, for $m_\chi \approx 100$ GeV, we expect the direct detection cross sections of order $\sigma \sim G_{\rm wk}^2 {m_\chi}^2 \sim 10^{-34}$ cm^2, where as the current direct detection experiments limits are $\sigma \lesssim 10^{-46}$ cm^2 [31] at $m_\chi \sim 100$ GeV. Thus the original *WIMP miracle* scenario as a dark matter model is ruled out and we must look for other more complex models of dark matter.

Chapter 3

Radiative transfer

In order to understand observations and connect them with theory, in particular the observations of the cosmic microwave background we need some tools and results from the theory of radiative transfer i.e. the interaction of photons with the medium they are passing through.

3.1 Radiative transfer

When dealing with light rays, it is convenient to recast the Boltzmann equation in terms of intensity, I_ν, to obtain the radiative transfer equation,

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\alpha_\nu I_\nu + j_\nu, \tag{3.1}$$

where, s is the distance along the path travelled by the light ray with $\mathrm{d}s = c\mathrm{d}t$, and as earlier, we have creation and destruction terms on the write hand side. The first term on the right hand side is absorption or scattering term, $\alpha_\nu = n\sigma_\nu$ is the coefficient of absorption, n is the number density of absorbers, σ_ν is the absorption cross section, and j_ν is the emission coefficient of photons by the medium along the direction of light ray. The differential optical depth is given by $\mathrm{d}\tau = n\sigma_\nu \mathrm{d}s = n\sigma_\nu c\mathrm{d}t$ and the optical depth, τ, along a light geodesic parameterized by distance s is

$$\tau = \int_s n\sigma_\nu \mathrm{d}s. \tag{3.2}$$

We can do a change of variable from s to τ in the radiative transfer equation giving

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau} = -I_\nu + J_\nu, \tag{3.3}$$

where $J_\nu = j_\nu/\alpha_\nu$ is the source function defined as the ratio of emission coefficient to absorption coefficient.

In the absence of any emission, the radiative transfer equation simplifies to

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau} = -I_\nu, \tag{3.4}$$

with solution

$$I_\nu = I_{\nu 0}e^{-\tau}. \tag{3.5}$$

Thus the intensity decreases exponentially with optical depth. We should note that the emission and absorption terms include scattering processes also. For example, scattering of photons out of the light ray we are following (or line of sight) would contribute to the absorption term, σ_ν, and scattering of photons from other direction into our line of sight would contribute to the emission term, J_ν.

We see from Eq. 3.5 that a fraction $e^{-\tau}$ photons survive absorbtion on traveling a distance τ. Thus, the probability that any photon will survive traveling a distance τ is $e^{-\tau}$ and that it will get absorbed before traveling a distance of τ is $1 - e^{-\tau}$. The mean distance in units of τ traveled by a photons is therefore,

$$\langle\tau\rangle = \int_0^\infty \tau e^{-\tau}\mathrm{d}\tau = 1 \tag{3.6}$$

If λ_{mfp} is the mean free path of the photon, it is related to the mean optical depth by $n\sigma_\nu\lambda_{\mathrm{mfp}} = \langle\tau\rangle = 1$ or

$$\lambda_{\mathrm{mfp}} = \frac{1}{n\sigma_\nu}. \tag{3.7}$$

Usually the optical depth varies along the photon path. We can define the differential optical depth $\dot{\tau}$ such that the total optical depth between a early time t and time when the photon is observed t_0 is given by

$$\begin{aligned}
\tau &= \int_t^{t_0} n(t')\sigma_\nu c\mathrm{d}t' \\
&= -\int_t^{t_0} \dot{\tau}(t')\mathrm{d}t' \\
\dot{\tau} &= -n\sigma_\nu c
\end{aligned} \tag{3.8}$$

Note that as t increases, the optical depth between t and the fixed observation time t_0 decreases giving the minus sign in definition of $\dot{\tau}$. The optical depth, $\Delta\tau = -\dot{\tau}\Delta t$ gives the average number of scatterings a photon suffers in time Δt or the probability of scattering in small time interval Δt. This can be seen from the fact that for small τ the probability of absorption is $1 - e^{-\tau} \approx \tau$. We can thus write the probability that a photon suffered a scattering at time t in the time interval Δt and did not suffer any scattering after that as

$$\begin{aligned}&\text{probability of scattering in }\Delta\text{t} \times \text{probability of survival after t}\\&= \dot{\tau} e^{-\tau} \Delta t\end{aligned} \tag{3.9}$$

where t_0 is the time when the photon is observed and τ is the optical depth traversed by the photon between time t and when it is observed. The quantity

$$g(t) = \dot{\tau} e^{-\tau} \tag{3.10}$$

is called the visibility function. When the scattering process is Thomson scattering, the visibility function defines the last scattering surface of the CMB.

Kirchoff's law

In thermal equilibrium absorption and emission must balance and radiation should have the blackbody spectrum (B_ν),

$$\begin{aligned}I_\nu &= B_\nu\\j_\nu &= \alpha_\nu B_\nu.\end{aligned} \tag{3.11}$$

This is known as Korchoff's law and relates the emission and absorption coefficients for any medium.

3.2 Line profile

We will be interested in electronic transitions between different atomic levels during recombination. Heisenberg's uncertainty principle tells us that there should be some quantum uncertainty in the frequency of the photon that is emitted in a transition from one atomic level to another, i.e.

$$\Delta E \Delta t \gtrsim \hbar, \tag{3.12}$$

where ΔE is the energy of the emitted photon. For example, the energy of a Lyman-α photon in transition from first excited level of hydrogen atom to ground level will not be exactly 10.2 eV. The only time scale in the problem is the lifetime of the atom on the excited state or the inverse of the transition rate given by the spontaneous emission coefficient A_{21} for transition between two levels labeled 1 and 2. It is also known as the Einstein A coefficient. The uncertainty in time, i.e. when the atom will decay, is of order $1/A_{21} \equiv 1/\gamma$.

Clasically, in terms of the electric field of the emitted photon, we have the Fourier relation between emitted frequency ω (with the energy of photon $E = \hbar\omega$) and times t,

$$\hat{E}(\omega) = \frac{1}{2\pi}\int E(t)e^{i\omega t} \tag{3.13}$$

If a time domain signal is of very short duration, it is extended in Fourier domain and vice versa. For example, an infinite sinusoid in time domain is a Dirac delta distribution in Fourier domain. For decay, the electric field is an exponentially decaying sinusoid, $e^{-i\omega_0 t}e^{-\gamma t/2}$, where $\omega_0 = 2\pi\nu_0$ is the resonance or rest frame line frequency and $\omega = 2\pi\nu$ is the angular frequency. Its Fourier transform[1], called Lorentzian profile, gives the spectrum of the emitted radiation. In quantum mechanical model of atom, the probability of finding an atom in excited state decreases exponentially as $e^{-\gamma t}$ and the wave function of the excited state has the time dependence $e^{-\gamma t/2}$ and we recover the classical result for the spectrum. The *naturally broadened* spectrum is given by the Lorentzian profile

$$\phi(\nu) = \frac{\gamma/(4\pi^2)}{(\nu-\nu_0)^2 + (\gamma/(4\pi))^2}. \tag{3.14}$$

In addition to the natural broadening of the line, the atoms emitting or absorbing the photons are moving with thermal velocities for a medium at a finite temperature T. Thus there will be a Doppler shift or Doppler broadening of the line. For a thermal Maxwell-Boltzmann (Gaussian) distribution of atoms, $\propto e^{-mv_x^2/(2k_\mathrm{B}T)}$, the Doppler shift is v_x/c, where v_x is the velocity along the line of sight which we take to be the x-axes and m is the mass of the atom. For rest frame freuency ν_0, shift is $\nu - \nu_0 = \nu_0 v_x/c$. The probability of atoms having veclocity between v_x and $v_x + \mathrm{d}v_x$ is $e^{-mv_x^2/(2k_\mathrm{B}T)}\mathrm{d}v_x$. We can do a change of variable from v_x to ν to get the probability of observing a photon at ν, $\phi(\nu) \propto e^{-mc^2(\nu-\nu_0)^2/(2\nu_0^2 k_\mathrm{B}T)}$. Normalizing the distribution gives the Doppler broadened line profile,

$$\phi(\nu) = \frac{1}{\Delta\nu_\mathrm{D}\sqrt{\pi}}e^{-(\nu-\nu_0)^2/\Delta\nu_\mathrm{D}^2}, \tag{3.15}$$

where the Doppler width, $\Delta\nu_\mathrm{D}$, is given by

$$\Delta\nu_\mathrm{D} = \nu_0\left(\frac{2k_\mathrm{B}T}{mc^2}\right)^{1/2} \tag{3.16}$$

[1]Note that the mathematica is not able to handle this Fourier transform and gives a dumb result after assuming that the decay rate γ is imaginary!

In general both effects can be present and we must combine the Lorentzian and Doppler profile to get what is known as the Voigt profile. There is no simple formula for the Voigt profile. To get the Voigt profile, we take the Lorentz profile and integrate it over the Doppler boosts to get an intergal formula for the Voigt profile.

3.3 Einstein Coefficients

The emission rate or transition probability per unit time from transition from an excited state 2 to a lower state 1 is called Einstein A coefficient. We will denote it by A_{21}. Similarly the absorption coefficient is B_{12} such that $\bar{I}B_{12}$ is the absorption rate or transition probability per unit time from state 1 to 2, where $\bar{I}$ is the average intensity over the line profile, $\phi(\nu)$. The line profile is normalized so that

$$\int_0^\infty \phi(\nu)\mathrm{d}\nu = 1$$
$$\bar{I} = \int I_\nu \phi(\nu)\mathrm{d}\nu. \tag{3.17}$$

We also have the stimulated emission coefficient, B_{21}, such that $\bar{I}B_{21}$ is the stimulated transition rate from the excited state 2 to the lower state 1.

The three Einstein coefficients are related to each other and we can find these relations using detailed balance as before. In thermodynamic equilibrium the emission and absorption must be equal. If n_i is the number of atoms in state i, we therefore have in equilibrium,

$$n_1 B_{12}\bar{I} = n_2\left(A_{21} + B_{21}\bar{I}\right). \tag{3.18}$$

Solving for the intensity we get,

$$\bar{I} = \frac{A_{21}/B_{21}}{n_1/n_2\,(B_{12}/B_{21}) - 1}. \tag{3.19}$$

In equilbrium at temperature T, we should have the populations of levels related by the Boltzmann factors,

$$\frac{n_1}{n_2} = \frac{g_1}{g_2}\frac{e^{-E_1/(k_\mathrm{B}T)}}{e^{-(E_1+h\nu_0)/(k_\mathrm{B}T)}}, \tag{3.20}$$

where E_1 is the energy of level 1, g_i is the degeneracy of the level i, and $h\nu_0$ is the energy difference between the two levels. Therefore, the solution for $\bar{I}$ becomes

$$\bar{I} = \frac{A_{21}/B_{21}}{g_1/g_2\,(B_{12}/B_{21})\,e^{h\nu_0/(k_\mathrm{B}T)} - 1}. \tag{3.21}$$

In equilibrium, the radiation field is given by the blackbody spectrum (B_ν),

$$\bar{I} = B_\nu = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/(k_B T)} - 1}. \tag{3.22}$$

Comparison of equations 3.21 and Eq. 3.22 gives the relations between the Einstein coefficients,

$$\begin{aligned} A_{21} &= \frac{2h\nu^3}{c^2}B_{21} \\ g_1 B_{12} &= g_2 B_{21} \end{aligned} \tag{3.23}$$

Chapter 4

Cosmic Microwave Background: homogeneous Universe

4.1 Thomson scattering

After the epoch of primordial nucleosynthesis and electron-positron annihilation, the main interaction of photons is Thomson scattering on free electrons or elastic scattering of photons and electrons.

$$\gamma + e^- \to \gamma + e^- \tag{4.1}$$

Since we are dealing with energies much smaller compared to the electron mass, we can work in the non-relativistic limit. Let $P = (p, \boldsymbol{p}), Q = \gamma_{\rm e} m_{\rm e}(1, \boldsymbol{v})$ denote the 4-momentum of photon and electron respectively before scattering, where $\gamma_{\rm e} = (1 - v^2)^{-1/2}$ is the Lorentz factor for the electron, $m_{\rm e}$ is the electron mass, $\boldsymbol{v}$ the velocity of electron, and $\boldsymbol{p}$ the 3-momentum of photon. We will denote 3-vectors with bold symbols and their magnitudes with the same symbols in normal font, and unit vectors with hatted quantities (e.g. $\hat{\mathbf{p}}$). Let us denote the energy-momentum of photon and electron after the collision by primed quantities, $P', Q', \boldsymbol{p}'$ etc.

Exercise 14

Use energy momentum conservation, $P + Q = P' + Q'$ to show that the momentum of photon after collision is

$$p' = \frac{p\,(1 - v\mu)}{(1 - v\mu') + \frac{p}{\gamma_{\rm e} m_{\rm e}}\,(1 - \cos\theta)}, \tag{4.2}$$

where $\cos\theta = \hat{\mathbf{p}}.\hat{\mathbf{p}}'$, $\mu = \hat{\upsilon}.\hat{\mathbf{p}}$, and $\mu' = \hat{\upsilon}.\hat{\mathbf{p}}'$. Take the non-relativistic limit, $\upsilon \ll 1$ to show that in this limit,

$$p' = \frac{p\,(1-\upsilon\mu)}{(1-\upsilon\mu') + \frac{p}{m_\mathrm{e}}\,(1-\cos\theta)}, \tag{4.3}$$

If the electron is initially at rest, ($\upsilon = 0$) we get,

$$p' = \frac{p}{1+\frac{p}{m_\mathrm{e}}(1-\cos\theta)} \tag{4.4}$$

and the change in momentum in the non-relativistic limit ($p/m_\mathrm{e} \ll 1$) is

$$\frac{\Delta p}{p} = \frac{p'-p}{p} \approx -\frac{p}{m_\mathrm{e}}\,(1-\cos\theta)\,. \tag{4.5}$$

Thus the energy transfer from hotter photons to colder electrons is small but non-zero. This small energy transfer is important to keep the electrons at the same temperature as photons until $z \approx 500$. The electrons, ions and atoms are also kept at the same temperature through coulomb interactions. In the absence of Thomson scattering, the electron-baryon plasma would cool much faster compared to the photons.

We had earlier derived that the momentum of every particle redshifts as $p \propto 1/a$. For electrons and baryons with Maxwell-Boltzmann momentum ditribution, $f(p) \propto e^{-p^2/(2mT)}$, conservation of phase space density implies $T \propto 1/a^2$ whereas the photon in isolation would cool as $T \propto 1/a$. Thus baryons are cooling faster compared to the photons and Thomson scattering must continuously transfer energy from photons to baryons to keep them at the same temperature. We can calculate approximately how much energy the photons lose to the baryons using conservation of entropy. Even though the entropy of photons and baryons is individually not conserved, the total entropy is approximately conserved.

It is convenient to work with entropy per baryon or s/n_B, where s is the entropy density and n_B is the baryon number density. Since both entropy and baryon number are conserved, s/n_B is also conserved. We again start with the second law of thermodynamics, Eq. 2.18 and use the fact that Entropy $S = sV$, thermal energy $E = \rho^\mathrm{th}V$ and volume $V \propto 1/n_\mathrm{B}$, where ρ^th is the total thermal energy density to get

$$\mathrm{d}(s/n_\mathrm{B}) = \frac{\mathrm{d}(\rho^\mathrm{th}/n_\mathrm{B}) + P\mathrm{d}(1/n_\mathrm{B})}{T} \tag{4.6}$$

For photons, $\rho^\mathrm{th}_\gamma = a_\mathrm{R}T^4, P_\gamma = 1/3a_\mathrm{R}T^4$ while for non relativistic particles, the internal energy density is given by $\rho^\mathrm{th}_\mathrm{B} = 3/2n_\mathrm{B}g_\mathrm{eb}k_\mathrm{B}T$ and pressure is given by the ideal gas law, $P_\mathrm{B} = n_\mathrm{B} g_\mathrm{eb} k_\mathrm{B} T$, where g_eb is the number of non-relativistc particles per baryon. For fully ionized plasma with 24% helium by mass, taking into account one electron per hydrogen and two electrons per helium atom,

$$g_\mathrm{eb} = \frac{2n_\mathrm{H} + 3n_\mathrm{He}}{n_\mathrm{B}} = \frac{0.76(2\rho_\mathrm{b}/m_\mathrm{p}) + 0.24(3/4)\rho_\mathrm{b}/m_\mathrm{p}}{\rho_\mathrm{b}/m_\mathrm{p}} \\ = 1.7, \tag{4.7}$$

where n_H is the number density of hydrogen atoms and n_He is the number density of helium atoms. We want to integrate Eq. 4.6 to get the entropy per baryon. Note that Eq. 4.6 is a total derivatve and in particular we must integrate over the two variables on the right hand side, n_B and T simultaneously. We therefore need to convert the right hand side also into a total derivative. Integrating over the baryon contribution is easy since it reduces trivially to sum of two terms each only a function of either n_B or T,

$$\left.\frac{s}{n_\mathrm{B}}\right|_\mathrm{B} = \int \frac{(3/2 g_\mathrm{eb} k_\mathrm{B})\mathrm{d}T}{T} + \int g_\mathrm{eb} k_\mathrm{B} n_\mathrm{B} \mathrm{d}(1/n_\mathrm{B}) \\ = g_\mathrm{eb} k_\mathrm{B} \ln\left(\frac{T^{3/2}}{n_\mathrm{B} C}\right), \tag{4.8}$$

where C is the (dimensionful) constant of integration which is not important for this calculation since we are not interested in absolute value of entropy.

Exercise 15

We can also derive entropy from free energy, F, using the thermodynamic relation

$$S = -\left.\frac{\partial F}{\partial T}\right|_V \tag{4.9}$$

Starting with the free energy of a monoatomic ideal gas, derive the expression for entropy per baryon s/n_B. What is the value of constant C in Eq. 4.8 ?

For the photon contribution to entropy we have,

$$
\begin{aligned}
\left.\frac{s}{n_{\rm B}}\right|_{\gamma} &= \int \frac{1}{T}{\rm d}\left(\frac{a_{\rm R}T^4}{n_{\rm B}}\right) + \frac{1}{3}\int a_{\rm R}T^3{\rm d}\left(\frac{1}{n_{\rm B}}\right)\\
&= a_{\rm R}\int \frac{4T^2}{n_{\rm B}}{\rm d}T + a_{\rm R}\int T^3{\rm d}\left(\frac{1}{n_{\rm B}}\right) + \frac{1}{3}a_{\rm R}\int T^3{\rm d}\left(\frac{1}{n_{\rm B}}\right)\\
&= \frac{4}{3}a_{\rm R}\left[\int \frac{3T^2}{n_{\rm B}}{\rm d}T + \int T^3{\rm d}\left(\frac{1}{n_{\rm B}}\right)\right]\\
&= \frac{4}{3}a_{\rm R}\frac{T^3}{n_{\rm B}} \qquad (4.13)
\end{aligned}
$$

Thus we have for the total entropy per baryon,

$$
\frac{s}{n_{\rm B}} = \frac{4a_{\rm R}T^3}{3n_{\rm B}} + g_{\rm eb}k_{\rm B}\ln\left(\frac{T^{3/2}}{n_{\rm B}C}\right). \qquad (4.14)
$$

The first term is proportional to the number density of photons to baryons, $1/\eta = n_\gamma/n_{\rm B} \sim 10^9$, where the second term is of order 1. The first term, the entropy of photons, therefore dominates the total entropy density by far. We note that $n_{\rm B} \propto 1/a^3$, since the total number of baryons (neutrons+protons) is conserved at $T \ll m_{\rm p}$. Conservation of entropy (or entropy per baryon) therefore implies that in the absence of any interactions between the photons and baryons, $T^3 \propto n_{\rm B}$ or $T \propto 1/a$. Similarly, for baryons, $T^{3/2} \propto n_{\rm B}$ or $T \propto 1/a^2$. Thus baryons cool at a rate that is twice as fast compared to the photons. Starting at some initial redshift z_i and temperature T_i, lets define $T_z = T_i(1+z)/(1+z_i)$. The actual *temperature* $T \equiv \left(\rho_\gamma/a_{\rm R}\right)^{1/4}$ of photons will depart from this relation because of energy transfer to baryons which are cooling faster. This energy transfer will keep the baryons at approximately the same temperature as the photons. We can parametrize this energy transfer as $T = T_z(1+t), t \ll 1$, where $t = (1/4)\Delta\rho_\gamma/\rho_\gamma$ is the fractional change in photon temperature. Substituting T in Eq. 4.14 and doing a Taylor series expansion in t keeping terms up to first order in t, we get

$$
\begin{aligned}
\frac{s}{n_{\rm B}}(z) &= \frac{4a_{\rm R}T_z^3(1+3t)}{3n_{\rm B}} + g_{\rm eb}k_{\rm B}\ln\left(\frac{T_z^{3/2}}{n_{\rm B}C}\right) + \frac{3}{2}g_{\rm eb}k_{\rm B}t\\
&= \frac{s}{n_{\rm B}}(z_i) \qquad (4.15)
\end{aligned}
$$

Solving for t we get

$$\begin{aligned}\frac{\Delta\rho_\gamma}{\rho_\gamma} = 4t &= -\frac{3g_{\rm eb}n_{\rm B}k_{\rm B}}{2a_{\rm R}T_z^3}\ln\left(\frac{1+z_i}{1+z}\right)\\ &= -\frac{\rho_{\rm b}^{\rm th}}{\rho_\gamma}\ln\left(\frac{1+z_i}{1+z}\right)\\ &= -5.9\times10^{-10}\ln\left(\frac{1+z_i}{1+z}\right)\end{aligned} \tag{4.16}$$

Thus, as expected, photons have to transfer a negligile amount of energy to keep the baryons at same temperature as them. This approximate equilibrium persists until the eneregy transfer rate from photons to baryons falls below the Hubble rate.

Exercise 16

The fractional energy transfer in each collision between the photons and electrons is given by (Eq. 4.5) $\approx p/m_{\rm e} \approx T/m_{\rm e}$. The collision rate of electrons with photons is $\approx n_\gamma\sigma_{\rm T}$, where $\sigma_{\rm T} = 6.65\times10^{-25}$ cm^2 is the Thomson scattering cross section and n_γ is the photon number density. Thus the energy transfer rate from photons to baryons is $\Gamma = Tn_\gamma\sigma_{\rm T}/m_{\rm e}$. After recombination, most of the electrons are bound up in neutral atoms, and only a small fraction of electrons, $x_{\rm e} = n_{\rm e}/n_{\rm B} \sim 10^{-4}$, where $n_{\rm e}$ is the number density of free electrons, remain free to transfer energy between the photons and baryons. Thus the actual energy transfer rate must be corrected for this, since energy extracted by each electron must be shared between $\sim 1/x_{\rm e}$ particles, and is given by $\Gamma x_{\rm e}$. Upto which redshift can this energy transfer through Thomson scattering keep the electrons at same temperature as photons ? Assume matter dominated Universe with $\Omega_{\rm m}h^2 = 0.14$. What would be the redshift up to which equilibrium is maintained if there was no recombination ?

We find in the above exercise that even the small amount of residual electrons after recombination, through Thomson scattering are able to keep the baryon temperature same as the CMB temperature until very late redshifts, $z \sim 400$. Thus even though *photons decouple from baryons* at $z \approx 1100$, the *baryons remain coupled to photons* until $z \approx 400$. Since baryons are taking away energy from the initially blackbody spectrum of the CMB, the spectrum of the CMB will get distorted. At redshifts $z \gtrsim 2\times10^6$, it is impossible to create deviations from the blackbody, since Thomson/Compton scattering in combination with photon creation and destruction processes of double Compton scattering and bremsstrahlung maintain equilibrium spectrum of photons. The total fractional energy taken from CMB between

$z \approx 2\times10^6$ and $z = 400$ is given by (Eq. 4.16) $5.9\times10^{-10}\ln(2\times10^6/400) = 5\times10^{-9}$. Thus we expect fractional deviations from a blackbody spectrum of order 10^{-9} due cooling of the CMB by baryons.

For most calculations, we can ignore this small change in the spectrum of the CMB and it is sufficient to take into account the fact that the baryon temperature remains the same as the CMB temperature and redshifts as $T \propto 1/a$, same as the CMB. Thus, to lowesr order, Thomson scattering changes the direction of photons and maintains the baryons at the same temperature as photons (without affecting the photons themselves significantly). The Thomson optical depth, or the scattering rate of photons with free electrons, $\tau_{\rm T} = \int n_{\rm e}\sigma_{\rm T}c{\rm d}t$, is quite large before recombination. In other words, the mean free path of a photon is very small compared to the horizon size, with a photon sufferring many scatterings in a Hubble time. As the universe expands, the scattering rate decreases, first slowly as a^{-3} due to decrease in electron density due to the expansion of the Universe. At $z \sim 1100$, the Universe undergoes a phase transition with the electrons combining with protons to form hydrogen atoms and the free electron density (hence the Thomson scattering rate) drops suddenly. The mean free path of the photons becomes suddenly larger than the horizon size, i.e. the photons after the hydrogen recombination free stream without suffering any significant scatterings. There is a small increase in the scattering rate when the first stars form at $z \sim 10-30$, and the energetic photons emitted by the first galaxies start reionizing the Universe.

The Thomson scattering of CMB photons, just by changing the direction of photon in the scattering, modifies the anisotropies or the primordial temperature fluctuations of the CMB and also creates polarization of the CMB photons. Thus, in order to understand the CMB temperature and polarization anisotropies and their power spectrum, we must first understand the process of recombination.

4.2 Recombination

Electrons remain in kinetic equilbrium with the atoms and ions through Coulomb interactions. The atoms and ions are thus also kinetically coupled to the photons with electrons acting as mediators. The scattering rate of *photons with electrons* is given by $n_{\rm e}\sigma_{\rm T}c$, where $n_{\rm e}$ is the electron number density and $\sigma_{\rm T} = 6.65\times10^{-25}$ cm^2. For a fully ionized primordial plasma with 24% helium mass fraction,

$$n_{\rm e} = n_{\rm H} + 2n_{\rm He} \approx (0.76 + 0.24/2)\frac{\rho_{\rm cr}\Omega_{\rm b}(1+z)^3}{m_{\rm p}}. \tag{4.18}$$

For $m_\mathrm{p} \approx 1.67 \times 10^{-24}$ g, $\rho_\mathrm{cr} \approx 8.5 \times 10^{-30}$ g/cm^3, $\Omega_\mathrm{b} \approx 0.049$, we have

$$n_\mathrm{e} \approx 189 \left(\frac{1+z}{1000}\right)^3 \ \mathrm{cm}^{-3}/ \tag{4.19}$$

The mean free path of photons λ_mfp is,

$$\lambda_\mathrm{mfp} = \frac{1}{n_\mathrm{e}\sigma_\mathrm{T}} \approx 8 \times 10^{21} \left(\frac{1000}{1+z}\right)^3 \ \mathrm{cm} = 2.6 \left(\frac{1000}{1+z}\right)^3 \ \mathrm{kpc} \tag{4.20}$$

This is very small compared to the horizon size at recombination ($z \approx 1000$), $\eta_* \approx 300$ Mpc. On scales $\lambda \gg \lambda_\mathrm{mfp}$, a fluid approximation is valid and we can assume that electron-baryons-photons behave as a single fluid. This is also known as the *tight coupling approximation*.

A photon, before recombination, is therefore scattered numerous times within a Hubble time. We can calculate the average number of scatterings or the optical depth of the photon for Thomson scattering, from time t until today (t_0), in radiation dominated Universe with $H(z) = H_0 {\Omega_\mathrm{r}}^{1/2}(1+z)^2$,

$$\tau_\mathrm{T}(z) = \int_t^{t_0} n_\mathrm{e}\sigma_\mathrm{T} c \mathrm{d}t' = \int_0^z \frac{n_\mathrm{e}\sigma_\mathrm{T} c}{H(1+z')}\mathrm{d}z' \approx 0.21 z. \tag{4.21}$$

During recombination, we are approximately in the matter dominated regime with $H(z) \approx H_0 {\Omega_\mathrm{m}}^{1/2}(1+z)^{3/2}$ and

$$\tau_\mathrm{T}(z) \approx 88 \left(\frac{1+z}{1100}\right)^{3/2}. \tag{4.22}$$

Thus the optical depth is about ~ 100 before recombination and increases with redshift. Note that the scattering rate of *electrons with photons* $\approx \tau_\mathrm{T} n_\gamma / n_\mathrm{e} \approx 10^9 \tau_\mathrm{T}$, is a billion times higher.

We will now ignore helium for simplicity and study the recombination of hydrogen. Since helium has higher ionization threshold, by the time hydrogen recombination starts, helium is almost fully recombined. Also, most of the electrons are contributed by the hydrogen even in the fully ionized plasma.

When the electrons finally combine with protons to form hydrogen atoms is decided by the competition between *recombination*

$$\mathrm{H}^+ + \mathrm{e}^- \rightarrow \mathrm{H} + \gamma \tag{4.23}$$

and photoionization

$$\mathrm{H} + \gamma \rightarrow \mathrm{H}^+ + \mathrm{e}^- \tag{4.24}$$

4.2.1 Recombination in equilibrium: Saha equation

Lets first assume that the recombination happens in equilibrium. If this were the case, we can use the Saha equation,

$$\frac{n_{\rm p}n_{\rm e}}{n_{\rm H}n_{\gamma}} = \frac{n_{\rm p}^{\rm eq}{n_{\rm e}}^{\rm eq}}{n_{\rm H}^{\rm eq}n_{\gamma}^{\rm eq}}, \tag{4.25}$$

where the protons, electrons and neutral hydrogen atom number densities are labeled with subscripts p, e, and H respectively and the equilibrium number densities of massive particles with mass m and degeneracy g at temperature T is given by

$$\begin{aligned} n_{\rm p,e,H}^{\rm eq} &= g\int \frac{{\rm d}^3p}{(2\pi\hbar)^3}e^{-E/(k_{\rm B}T)} \\ &= g\left(\frac{mk_{\rm B}T}{2\pi\hbar^2}\right)^{3/2} e^{-m/(k_{\rm B}T)}. \end{aligned} \tag{4.26}$$

Since we have assumed that the photons have blackbody spectrum, $n_{\gamma} - n_{\gamma}^{\rm eq}$. In reality, this is the assumption that will fail and makes our Saha solution invalid. In particular, the photon emitted during the recombination process make the CMB spectrum depart from the blackbody spectrum and it is extremely important to take these extra photons into account.

The Saha equation under the assumption that CMB has a blackbody spectrum becomes,

$$\frac{n_{\rm p}n_{\rm e}}{n_{\rm H}} = \frac{g_{\rm p}g_{\rm e}}{g_{\rm H}}\left(\frac{m_{\rm e}k_{\rm B}T}{2\pi\hbar^2}\right)^{3/2} e^{-(m_{\rm p}+m_{\rm e}-m_{\rm H})/(k_{\rm B}T)} \tag{4.27}$$

Exercise 17

Calculate the degeneracy factor of hydrogen atom, $g_{\rm H}$?

The sum of mass of electron and proton differ from the mass of hydrogen by the binding energy (or ionization energy) of the hydrogen atom, $E_{\rm I} = m_{\rm p} + m_{\rm e} - m_{\rm H} =$ 13.6 eV. The degeneracy factors are $g_{\rm e} = g_{\rm p} = 2, g_{\rm H} = 4$. Lets also define ionization fraction, $x_{\rm e}$ as the ratio of free electrons to total hydrogen (ionized as well as neutral), $n_{\rm H+p} = n_{\rm H} + n_{\rm p}$,

$$x_{\rm e} = \frac{n_{\rm e}}{n_{\rm p} + n_{\rm H}}. \tag{4.28}$$

Since we are ignoring helium, $n_\mathrm{p} = n_\mathrm{e}$ and we get

$$\begin{aligned}
\frac{n_{\mathrm{H+p}}{x_\mathrm{e}}^2}{1-x_\mathrm{e}} &= \left(\frac{m_\mathrm{e}k_\mathrm{B}T}{2\pi\hbar^2}\right)^{3/2} e^{-E_\mathrm{I}/(k_\mathrm{B}T)}\\
\frac{{x_\mathrm{e}}^2}{1-x_\mathrm{e}} &= \frac{1}{n_{\mathrm{H+p}}(0)(1+z)^3}\left(\frac{m_\mathrm{e}k_\mathrm{B}T}{2\pi\hbar^2}\right)^{3/2} e^{-E_\mathrm{I}/(k_\mathrm{B}T)},
\end{aligned} \tag{4.29}$$

where $n_{\mathrm{H+p}}(0) = \rho_\mathrm{cr}\Omega_\mathrm{b}/m_\mathrm{H} \approx 1.9\times10^{-7}\ \mathrm{cm}^{-3}$ is the total hydrogen number density today. At high temperatures, $T \to \infty$, $x_\mathrm{e} \to 1$ and as $T \to 0$, the solution is $x_\mathrm{e} \to 0$ or fully recombined hydrogen. The CMB temperature is $T(z) = T_\mathrm{CMB}(1+z) = 2.725(1+z)\ \mathrm{K} = 2.35\times10^{-4}(1+z)\ \mathrm{eV}$. Putting in the numbers, the Saha solution is

$$\frac{{x_\mathrm{e}}^2}{1-x_\mathrm{e}} = \frac{5.7\times10^{22}}{(1+z)^{3/2}} e^{-5.8\times10^4/(1+z)}. \tag{4.30}$$

When the temperature of the Universe is of order E_I, at $z \approx 5.8\times10^4$, $1-x_\mathrm{e} \approx 10^{-15}$, i.e. a negligible fraction of hydrogen is in recombined state. This is, as before, because of the large photon to baryon ratio of the Universe. For recombination to start, we need the exponential in Eq. 4.29 to be of order 10^{-15}. This happens when

$$1+z = \frac{-5.8\times10^4}{\ln(10^{-15})} \approx 1679. \tag{4.31}$$

At $z \lesssim 1600$, according to the Saha equation, the ionization fraction will drop exponentially. As recombination starts, for $x_\mathrm{e} \ll 1$ we can approximate ${x_\mathrm{e}}^2/(1-x_\mathrm{e}) \approx {x_\mathrm{e}}^2$ giving the solution

$$\begin{aligned}
x_\mathrm{e} &\approx \frac{2.4\times10^{11}}{(1+z)^{3/4}} e^{-2.9\times10^4/(1+z)}\\
x_\mathrm{e}(z=1300) &= 0.24\\
x_\mathrm{e}(z=1200) &= 0.04\\
x_\mathrm{e}(z=1100) &= 5\times10^{-3}
\end{aligned} \tag{4.32}$$

Thus according to the Saha solution, most of the hydrogen is recombined by $z \approx 1200$.

4.2.2 Case B recombination

However the recombination does not proceed in equilibrium. There are two reasons which delay recombination compared to the prediction of the Saha equation. Every successful recombination will produce a photon with energy $E_\gamma \geq E_\mathrm{I} = 13.6$ eV. This photon cannot escape. The cross-section for the photo-absorption of this photon on a neutral hydrogen a is $\sigma_\nu = 6.3 \times 10^{-18}\ \mathrm{cm}^2 \approx 10^7 \sigma_\mathrm{T}$. The Thomson optical depth, $\tau_\mathrm{T} \gg 1$, at $z = 1000$ and the optical depth for the 13.6 eV photons is $\gg 10^7$. Thus any photon emitted in recombination of the hydrogen atom will be immediately absorbed by a previously recombined atom, even when only a tiny fraction of atoms have recombined, $1 - x_\mathrm{e} \sim 10^{-7}$, resulting in net zero change in the ionization fraction. Thus, recombinations directly to the ground state of hydrogen are ineffective and cosmological recombination must proceed in a two step process. First the electron recombines to an excited state of the hydrogen atom, H^*,

$$p + e \rightarrow \mathrm{H}^* + \gamma(E \ll 10.2\ \mathrm{eV}), \tag{4.33}$$

emitting a photon with energy of order the ionization energy of the first excited state of hydrogen atom, $E_{\mathrm{I},2} = 3.4$ eV or smaller. Then the atom deexcites with the electron transitioning to the ground state either directly or first going through one or more intermediate states,

$$\mathrm{H}^* \rightarrow \mathrm{H} + \gamma\,(E < 13.6\ \mathrm{eV}) + \gamma + \cdots, \tag{4.34}$$

where the highest energy photon emitted as the electron cascades down to the ground level will have energy less than the ionization energy from the ground state.

Since we do not have equilbrium evolution, we cannot use Saha equation and must evolve the Botzmann equation for the free electron number density,

$$\begin{aligned} \frac{\mathrm{d}n_\mathrm{e}}{\mathrm{d}t} &= \text{photoionization} - \text{recombination} \\ \frac{\partial n_\mathrm{e}}{\partial t} + 3Hn_\mathrm{e} &= \beta n_{1\mathrm{s}} - \alpha n_\mathrm{e} n_\mathrm{p}, \end{aligned} \tag{4.35}$$

where $n_{1\mathrm{s}}$ is the number density of recombined hydrogen atoms in the ground 1s state, and n_H includes all recombined atoms including those in excited states. Since recombination happens at a temperature much smaller compared to the excitation energy from ground state to the first excited state, $E_{12} = 10.2$ eV, and de-excitation is extremely fast compared to other processes, most of the recombined hydrogen atoms are in ground state and $n_\mathrm{H} \approx n_{1\mathrm{s}}$ to a very good approximation. It is convenient to use fractional abundances instead of number densities of different species,

$$x_\mathrm{e} \equiv \frac{n_\mathrm{e}}{n_\mathrm{H} + n_\mathrm{p}}, x_{1\mathrm{s}} \equiv \frac{n_{1\mathrm{s}}}{n_\mathrm{H} + n_\mathrm{p}}, x_\mathrm{p} \equiv \frac{n_\mathrm{p}}{n_\mathrm{H} + n_\mathrm{p}}. \tag{4.36}$$

Helium, with ionization energy of first electron at 24.6 eV and that of the second electron (ionization energy of He^+ ion) at 54.4 eV, recombines earlier than hydrogen. Once helium has recombined, the electrons are contributed solely by hydrogen and $n_e = n_p$ or $x_e = x_p$. Also, since x_e is the ratio of two number densities, it does not change due to the expansion of the Universe. The Boltzmann equation for x_e can be written as

$$\frac{dx_e}{dt} = \beta x_{1s} - \alpha {x_e}^2 \left(n_H + n_p\right). \tag{4.37}$$

We need the recombination and photoionization coefficients, which are functions of temperature T, to solve this equation. Lets first consider the recombination coefficient, $\alpha(T)$. The electron can recombine to any of the levels of hydrogen and de-excite to the ground state. We should sum over recombination rates to all excited levels of hydrogen to get the total recombination rate, omitting the ground state since direct recombination to the ground state creates a photon with energy > 13.6 eV which immediately ionizes another atom. This recombination coefficient is known as the case B recombination coefficient, α_B,

$$\alpha_B(T) = \sum_{n=2}^{\infty} \sum_{\ell=0}^{n-1} \alpha_{n\ell}(T) \tag{4.38}$$

where $\alpha_{n\ell}$ is the recombination coefficient for level with prinicipal quantum number n and orbital quantum number ℓ. If we include $n = 1$ level in the sum, it is called case A recombination. Since the electron after recombining to an excited state de-excites immediately to the ground state, we need only consider ionizations from the ground state. Thus $\beta = \beta_{1s}$. We can find β in terms of α_B using detailed balance as before, by requiring that in equilibrium we should recover the Saha result. Using $x_{1s} = 1 - x_e, n_e = n_p$, we have

$$\frac{dx_e}{dt} = \beta (1 - x_e) - \alpha_B {x_e}^2 \left(n_H + N_p\right). \tag{4.39}$$

In equilibrium, $dx_e/dt = 0$, and we get (using Eq. 4.29

$$\begin{aligned} \frac{{x_e}^2}{1 - x_e} &= \frac{\beta}{\alpha_B \left(n_H + N_p\right)} \\ &= \frac{1}{n_H + N_p} \left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-E_I/(k_B T)} \\ \beta &= \left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-E_I/(k_B T)} \alpha_B. \end{aligned} \tag{4.40}$$

We thus get the Boltzmann equation describing the hydrogen recombination,

$$\frac{\mathrm{d}x_\mathrm{e}}{\mathrm{d}t} = \alpha_B \left[(1 - x_\mathrm{e}) \left(\frac{m_\mathrm{e} k_\mathrm{B} T}{2\pi\hbar^2} \right)^{3/2} e^{-E_\mathrm{I}/(k_\mathrm{B}T)} - x_\mathrm{e}{}^2 \left(n_\mathrm{H} + N_\mathrm{p} \right) \right]. \tag{4.41}$$

As the Universe expands and cools, the first term becomes negligible as the temperature of the photons drops. The second term also becomes negligible (compared to the Hubble rate) as the number density of electrons and protons $\propto (1+z)^3$ decreases. Thus, first the recombination is delayed w.r.t the Saha solution because atoms cannot directly recombine to the ground state. However, before all the atoms can recombine, the recombination rate becomes negligible because of decrease in the number density of electrons and protons and the recombination freezes out with a small residual fraction of free electrons.

A fitting formula for the case B recombination coefficient has been provided in [32],

$$\begin{aligned} \alpha_B &= 10^{-13} \frac{a t^b}{1 + c t^d} \ \mathrm{cm}^3 \mathrm{s}^{-1} \\ a &= 4.309, b = -0.6166, c = 0.6703, d = 0.5300, t = \frac{T}{10^4 \ \mathrm{K}}. \end{aligned} \tag{4.42}$$

We can do a change of variables from time to redshift in the Eq. 4.41 to the competition between the recombination rate and Hubble rate explicit,

$$\begin{aligned} (1+z)\frac{\mathrm{d}x_\mathrm{e}}{\mathrm{d}z} &= \frac{\mathrm{d}x_\mathrm{e}}{\mathrm{d}\ln(1+z)} \\ &= -\frac{\alpha_B}{H(z)} \left[(1 - x_\mathrm{e}) \left(\frac{m_\mathrm{e} k_\mathrm{B} T}{2\pi\hbar^2} \right)^{3/2} e^{-E_\mathrm{I}/(k_\mathrm{B}T)} - x_\mathrm{e}{}^2 \left(n_\mathrm{H} + N_\mathrm{p} \right) \right]. \end{aligned} \tag{4.43}$$

As long as the recombination rate is much larger compared to the Hubble rate, we should have equilibrium and follow the Saha solution. The photons will decouple from the electrons once the Thomson optical depth drops below 1. We saw earlier, that the Thomson optical depth at $z \sim 1100$ is ~ 100 for a fully ionized plasma. Therefore, the Thomson optical depth will become of order unity when the ionization fraction becomes $x_\mathrm{e} \sim 0.01$. Lets compare the recombination rate with the Hubble rates at different redshifts assuming the Saha solution (Eq. 4.32).

z	x_e(Saha)	$H(z)$ (s^{-1})	$\alpha_B x_\mathrm{e}{}^2(n_\mathrm{H}+n_\mathrm{p})$ (s^{-1})
1300	0.24	6.7×10^{-13}	1.4×10^{-11}
1200	0.04	5.9×10^{-14}	3.3×10^{-13}
1100	5×10^{-3}	5.2×10^{-14}	4.2×10^{-15}

(4.44)

We see that at $z = 1300$, the recombination (and photoionization) rates are significantly faster compared to the Hubble rate. Therefore, Saha solution is a good approximation until $z = 1300$. However, by $z = 1200$, the two rates are becoming comparable and we should expect departures from the Saha solution, and the optical depth of $\tau_T \approx 1$ or $x_e \approx 0.01$ will be reached a little later compared to the prediction of the Saha solution. The case A recombination rate is about 45% bigger compared to the case B recombination rate at $T = 3270$ K [32]. Therefore, if the we had kept recombinations to the ground state also in our recombination rate the ionization fraction would have followed the equilibrium solution for a little longer and we would have reached $\tau_T = 1$ a little earlier.

We can find the approximate redshift when the solution departs from the equilibrium solution by using the condition that the Hubble rate is equal to the recombination rate or ionization rate (since both are same in the Saha solution) and using the fact that it happens when $x_e \ll 1$,

$$H = \alpha_B {x_e}^2(n_H + n_p) = \alpha_B \left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-E_I/(k_B T)} \tag{4.45}$$

Taking the logarithm and solving for redshift, and taking values corresponding to $z = 1200$ inside the logarithm we get for the *freezeout* redshift z_f, upto logarithmic corrections, with $T(z) = 2.725(1 + z)$ K,

$$1 + z = \frac{E_I}{k_B T(0)} \left[\ln\left[\frac{\alpha_B}{H}\left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2}\right]\right]^{-1} \approx 1160 \tag{4.46}$$

At $z_f < 1160$, the photoionization rate drops exponentially while the recombibation rate drops only as a power law. To find the final residual electron fraction, we can therefore ignore the photo-ionizations. The evolution equation in this case simplifies to

$$\frac{dx_e}{dz} = \frac{\alpha_B}{(1 + z)H(z)} {x_e}^2 \left(n_H + N_p\right) \tag{4.47}$$

We can integrate it from z_f to $z = 0$, assuming α_B is constant and equal to its value at z_f, $\alpha_B \approx 6.4 \times 10^{-13}\text{cm}^3\text{s}^{-1}$. This is a good approximation since we most of the contribution to the recombinations would come from high redshifts close to z_f. Assuming matter dominated Universe, $H(z) \approx H_0 {\Omega_m}^{1/2}(1 + z)^{3/2} \approx 1.2 \times 10^{-18}(1+z)^{3/2}$ s^{-1} and $n_H + n_p = (0.76\Omega_b\rho_{cr}/m_p)(1+z)^3 \approx 1.9\times10^{-7}(1+z)^3$ cm^{-3}. Substituting these values, the evolution equation for x_e is

$$\frac{dx_e}{dz} = 0.1{x_e}^2(1 + z)^{1/2} \tag{4.48}$$

with solution

$$\frac{1}{x_e(0)} - \frac{1}{x_e(z_f)} = \frac{2}{3}0.1(1+z_f)^{3/2}. \tag{4.49}$$

From Eq. 4.32, $x_e(z_f) \approx 0.018$. We expect residual ionization fraction to be much smaller than $x_e(z_f)$ and therefore ignore this term, giving $x_e(0) \approx 15/(1+z_f)^{3/4} = 3.8 \times 10^{-4}$.

4.2.3 Recombination with 3-level atom

A nice discussion of recombination in multi-level atom can be found in [33] and the connection to the effective 3-level atom in [34]. There are assumptions in the above calculation which are still incorrect. We correctly assumed that almost all recombinations will happen through excited states of the atom. However, recombinations to the excited states are still not all successful. Some of the electrons will end up in the $2s$ state of hydrogen and transitions from $2s$ to $1s$ state are forbidden (why?). The electrons which end up in $2p$, or in general np levels, will transition to $1s$ ground state immediately. However, these Lyman series transitions are too strong, i.e. the absorption cross-section for the photons emitted through np to $1s$ transitions are very large and these photons will be immediately absorbed by recombined atoms in ground state. The ionization energy from $n = 2$ level is just 3.4 eV, and there is high probability that these excited atoms will get ionized.

Even though an atom in $2s$ state cannot de-excite to $1s$ state by emitting a single photon due to conservation of angular momentum, it can de-excite by emitting two photons,

$$H^*(2s) \rightarrow H(1s) + 2\gamma, \tag{4.50}$$

such that the sum of energies of two photons adds up to the energy difference between the levels, $E_{\gamma 1} + E_{\gamma 2} = 10.2$ eV.

The transition rate for the two photon transition from $2s$ to $1s$ state is $A_{2s1s} = 8.22\ \mathrm{s}^{-1}$. This should be compared to the allowed transition from $2p$ to $1s$ with $A_{2p1s} = 6.27 \times 10^8\ \mathrm{s}^{-1}$. Even though the two photon transition is much slower, it is still much faster compared to the Hubble rate and there when the $2p$ transition is blocked because most of the Ly-alpha series photons cannot escape, the two photon $2s - 1s$ transition becomes important.

We can estimate what fraction of Lyman-α photons can escape by comparing the absorption rate with the expansion rate. The absorption coefficient for the Lyman-α photons is given by, using Einstein relations,

$$B_{1s2p} = \frac{g_{2p}}{g_{1s}} B_{2p1s} = \frac{3c^2}{2h\nu^3} A_{2p1s}. \tag{4.51}$$

The absorption probability per unit times is given by

$$B_{\rm 1s2p}\bar{I} = B_{\rm 1s2p}\int I_\nu \phi(\nu){\rm d}\nu \tag{4.52}$$

Therefore, $B_{\rm 1s2p}I_\nu\phi(\nu)$ is the absorption probability per unit frequency per unit time. If the absorption/radiation is isotropic, we have

$$n_{1s}\frac{h\nu}{4\pi}B_{\rm 1s2p}I_\nu\phi(\nu) = \frac{\text{energy absorbed}}{{\rm d}V.{\rm d}t.{\rm d}\nu.{\rm d}\Omega}, \tag{4.53}$$

where ${\rm d}V$ is the infinitesimal volume and ${\rm d}\Omega$ is the infinitesimal solid angle, n_{1s} is the number density in the $1s$ state.

We can compare the above equation with a similar expression using the absorption cross section σ_{1s2p},

$$n_{1s}\sigma_{1s2p}I_\nu = \frac{\text{energy absorbed}}{{\rm d}\Omega.{\rm d}t.{\rm d}V.{\rm d}\nu}. \tag{4.54}$$

Comparing expressions 4.53 and 4.54 gives

$$\sigma_{1s2p} = \frac{h\nu}{B_{\rm 1s2p}}\phi(\nu) = \frac{3c^2}{8\pi\nu^2}A_{\rm 2p1s}\phi(\nu). \tag{4.55}$$

Similarly, the emission coefficient is given by

$$j_\nu = \frac{h\nu}{4\pi}\phi(\nu)n_{2p}. \tag{4.56}$$

Ignoring the Doppler broadening, the line profile is given by the Lorentzian profile,

$$\phi(\nu) = \frac{A_{\rm 2p1s}/(4\pi^2)}{(\nu-\nu_0)^2 + (A_{\rm 2p1s}/(4\pi))^2}. \tag{4.57}$$

At the line center, $\nu = \nu_0 = 2.5\times 10^{15}\ {\rm s}^{-1}$, $\phi(\nu_0) = 4/A_{\rm 2p1s}$ and

$$\sigma_{1s2p}(\nu_0) = \frac{12c^2}{8\pi\nu_0^2} = 7\times 10^{-11}\ {\rm cm}^2. \tag{4.58}$$

This is huge compared to the Thomson cross section. Away from the line center, the cross section decreases but still remains substantial,

$$\nu = \nu_0 \pm 10^{-5}\nu_0 : \sigma_{1s2p} = 2.7\times 10^{-16}\ {\rm cm}^2$$
$$\nu = \nu_0 \pm 10^{-4}\nu_0 : \sigma_{1s2p} = 2.7\times 10^{-18}\ {\rm cm}^2$$

The optical depth when travelling a distance of order horizon size at recombination, $300/(1+z_*)$ Mpc, where $\eta_* = 300$ Mpc is approximately the comoving horizon at recombination, $z_* \approx 1000$, is

$$\begin{aligned}\tau_{\rm Ly-\alpha} &= n_{\rm H+p}(1-x_{\rm e})\sigma_{1s2p}\frac{\eta_*}{1000}\\ &= 10^8 - 10^{10}(1-x_{\rm e})\end{aligned} \tag{4.59}$$

for $\nu_0 - 10^{-4}\nu_0 \lesssim \nu \lesssim \nu_0 - 10^{-5}\nu_0$. This is huge and imply that most of the Ly-α photons will be immediately absorbed as soon as there is a tiny neutral fraction. A small number of photons however will be able to redshift out of the line due to the expansion of the Universe and escape. The successful recombinations correspond to these escaped photons.

Whether most recombinations pass through the $2s$ level or $2p$ level depends on the time an atom spends in a level, i.e. on the transition probability. Since there is competition between the ionization from the second level, which need a photon of energy only $\gtrsim$ 3.4 eV, and transition to the ground state, where the atom will be safe, the faster transition rate will make recombinations faster. Even though the $2p-1s$ transitions are many orders of magnitude faster compared to the $2s-1s$ transitions, the fact that the Ly-α photons cannot escape means that the much slower $2s-1s$ transition is important. It turns out that the escape probability for the Ly-α photons, $P_{\rm esc}$ is of order 10^{-8}. Thus the effective rate of recombinations through the $2p$ channel is reduced by this factor and becomes comparable to the recombination rate through the $2s$ channel. In fact $2s$ channel dominates with $\sim 57\%$ of the recombinations happenning thorough the slower $2s-1s$ two photon transition [35]. Note that the photons emitted in the blue wing of the Ly-α line will pass through the line center as they redshift and have no chance of escaping. It is only the photons in the red wing of the line $\nu < \nu_0$, which will have some chance of getting out of the resonance and escape.

The case B recombination makes the assumption that every recombination to the excited state is successful. We see now that this is not true for the recombinations that pass through np levels to the ground state. Also, as we go to higher and higher levels, it becomes exponentially easier for the atoms to be photoionized again since there are exponentially more photons as we move up the Wien tail of the blackbody spectrum. Thus most successful recombination will pass through the $n=2$ level, either by direct recombination to the higher levels or recombination to an excited state and then cascading down to $n=2$ level. We therefore need to at least resolve $n=2$ state, in addition to the ground state and the ionized state of hydrogen atom. This is known as the 3-level model of the hydrogen atom [36, 37]. As before we will work with the fractional abundances in different levels w.r.t. the total hydrogen number density, $x_i = n_i/(n_{\rm H}+n_{\rm p})$.

Let f_{2s} be the fraction of recombinations that pass trhough the $2s$ level and $1 - f_{2s}$ the fraction that pass through the $2p$ level. Let $R_{\rm 2s1s}$ be the net transition rate from $2s$ to $1s$ level through two photon transition, after taking into account stimulated emission and absorption, i.e.

$$n_{2s}R_{\rm 2s1s} = A_{\rm 2s1s}n_{2s} - A_{\rm 1s2s}n_{1s}, \tag{4.60}$$

where $A_{\rm 1s2s}$ includes stimulated emission and absorption. We can find $A_{\rm 1s2s}$ in terms of $A_{\rm 2s1s}$ using detailed balance. In equilibrium the net transition rate must vanish and populations should be also be in Saha equilbrium giving,

$$A_{\rm 1s2s} = \frac{n_{2s}}{n_{1s}}A_{\rm 2s1s} = A_{\rm 2s1s}e^{-h\nu_\alpha/(k_{\rm B}T)}, \tag{4.61}$$

where ν_α is the Lyman-α transition frequency with $h\nu_\alpha = 10.2$ eV. Since recombination is happening at $k_{\rm B}T \ll h\nu_\alpha$, $A_{\rm 1s2s}$ is negligible and $R_{\rm 2s1s} \approx A_{\rm 2s1s}$. The population if the $2s$ level is decided by the fraction of total recombinations to the excited states which pass through $2s$ level minus the ionizations from the $2s$ level and deexcitation to the $1s$ level. We therefore have the following evolution equation for the $2s$ state,

$$\begin{aligned}\frac{{\rm d}x_{2s}}{{\rm d}t} &= \text{ recombinations } - \text{ ionizations } - \text{ net deexcitations}\\ &= f_{2s}\alpha_B {x_{\rm e}}^2 n_{\rm H+p} - \beta_{2s}x_{2s} - R_{\rm 2s1s}x_{2s}\\ &= f_{2s}\alpha_B {x_{\rm e}}^2 n_{\rm H+p} - \beta_{2s}x_{2s} - A_{\rm 2s1s}\left(x_{2s} - x_{\rm 1s}e^{-h\nu_\alpha/(k_{\rm B}T)}\right),\end{aligned} \tag{4.62}$$

where β_i is the ionization rate from level i. For the $2p$ state, we need to take into account escape probability, $P_{\rm esc}$, and the net transition rate from $2p$ to $1s$ is given by $P_{\rm esc}R_{\rm 2p1s}$, where

$$\begin{aligned}x_{2p}R_{\rm 2p1s} &= A_{\rm 2p1s}x_{2p} + B_{\rm 2p1s}B_\nu(T)x_{2p} - x_{\rm 1s}B_{\rm 1s2p}B_\nu(T)\\ &= A_{\rm 2p1s}\left[x_{2p}\left(1 + e^{-h\nu_\alpha/(k_{\rm B}T)}\right) - 3x_{\rm 1s}e^{-h\nu_\alpha/(k_{\rm B}T)}\right]\\ &\approx A_{\rm 2p1s}\left[x_{2p} - 3x_{\rm 1s}e^{-h\nu_\alpha/(k_{\rm B}T)}\right],\end{aligned} \tag{4.63}$$

where in the last line we ignore the small stimulated emission term ($e^{-h\nu_\alpha/(k_{\rm B}T)}$). We cannot neglect the other exponential term since x_{2p} can be much smaller compared to $x_{\rm 1s}$. We therefore have,

$$\begin{aligned}\frac{{\rm d}x_{2p}}{{\rm d}t} &= (1 - f_{2s})\alpha_B {x_{\rm e}}^2 n_{\rm H+p} - \beta_{2p}x_{2p} - P_{\rm esc}R_{\rm 2p1s}x_{2p}\\ &= (1 - f_{2s})\alpha_B {x_{\rm e}}^2 n_{\rm H+p} - \beta_{2p}x_{2p} - P_{\rm esc}A_{\rm 2p1s}\left[x_{2p} - 3x_{\rm 1s}e^{-h\nu_\alpha/(k_{\rm B}T)}\right].\end{aligned} \tag{4.64}$$

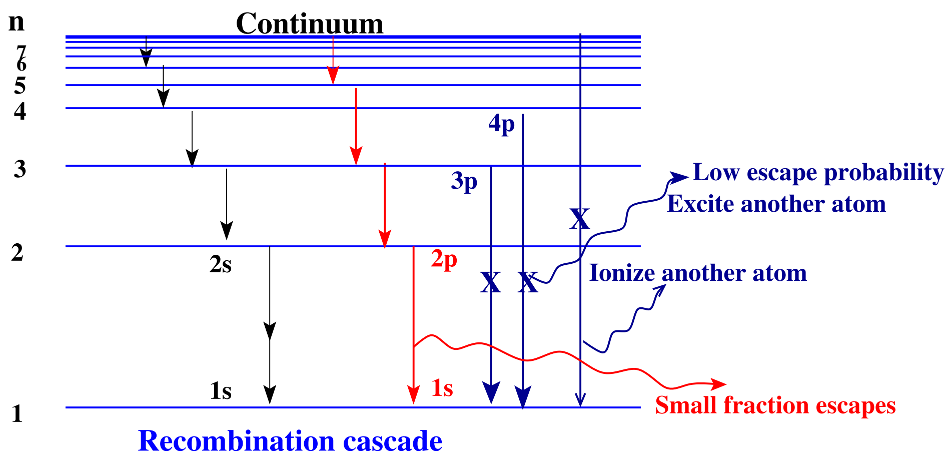

Figure 4.1

We note that we are still using the case B recombination coefficient, α_B, in the above equations. We are therefore making the assumption that the electron recombining to any higher level would eventually end up in $n = 2$ level before reaching the ground state, i.e., direct transitions from $n > 2$ levels to the ground state are neglected. This assumption is justified as follows. For an electron in any level $n\ell$ for $n > 2$, it has choice of transitioning to any level $n' < n$. The allowed transitions, in which ℓ changes by $\Delta\ell = \pm 1$, are much faster compared to other transitions such as two photon transitions. This means that the direct transitions to the $1s$ state are important only if the electron is in np state, otherwise the electron is much more likely to transition to other $n \geq 2$ states for which $\Delta\ell = \pm 1$. For the electron in np state we are again faced with the problem of escape of the Lyman-series photon. Even if the electron transitions from np to $1s$ state, the resulting Lyman series photon will be immediately absorbed. The extremely low escape probability of Lyman series photons, $P_{\rm esc} \ll 1$, imply that the net probability of successful transition of a np photon to $1s$, $P_{\rm esc}A_{np1s}$ is negligible compared to the the probability of transition to a ns state with $n' > 1$, $A_{npn's}$, since $n' > 1$ states have negligible populations and these photons escape with probability 1. The above discussion is schematically summarized in Fig. 4.1. Note that the electrons in $2p$ level also face the same problem of low escape probability. However, these electrons have no choice since $1s$ state is the only lower level available to them.

We can also find the ionization rates from $n = 2$ level in terms of the recombination coefficients using detailed balance. In equilibrium, all transitions must be individually balanced. In particular transitions between the atomic levels must be balanced giving $R_{\rm 2s1s} = R_{\rm 2p1s} = 0$ and photoionizations must be balanced by recombinations. Using equilbrium level populations, $n_i^{\rm eq}$ we get,

$$\begin{aligned}\alpha_B {n_{\rm e}}^{\rm eq} n_{\rm p}^{\rm eq} f_{2s} &= \beta_{2s} n_{2s}^{\rm eq}\\ \beta_{2s} &= f_{2s}\alpha_B \frac{{n_{\rm e}}^{\rm eq} n_{\rm p}^{\rm eq}}{n_{2s}^{\rm eq}}\\ &= f_{2s}\alpha_B \left(\frac{m_{\rm e}k_{\rm B}T}{2\pi\hbar^2}\right)^{3/2} e^{-E_{\rm I,2}/(k_{\rm B}T)}\\ &\equiv 4\beta_B f_{2s},\end{aligned} \tag{4.65}$$

where we have defined the case B ionization coefficient, β_B. Similarly for the $2p$ level we get

$$\beta_{2p} = (1-f_{2s})\alpha_B \frac{{n_{\rm e}}^{\rm eq} n_{\rm p}^{\rm eq}}{n_{2p}^{\rm eq}} = \frac{g_{2s}}{g_{2p}} 4\beta_B(1-f_{2s}) = \frac{1}{3}4\beta_B(1-f_{2s}) \tag{4.66}$$

since $n_{2p}^{\rm eq}/n_{2s}^{\rm eq} = g_{2p}/g_{2s} = 3$, where g_i is the degeneracy of level i.

In order to calculate what fraction of recombinations end up in $2s$, f_{2s}, we must solve for the higher levels also. However, the ionization energy of $2s$, $2p$ and higher levls is small and the number densities of photons which can photoionize from these levels is therefore large. The recombibnations and ionizations to the excited levels are therefore very fast compared to the Hubble rate and we can assume that the excited states are in thermal equilbrium relative to each other. In particular, the $2s$ and $2p$ levels, since they have the same energy are just filled according to their degeneracy factors in the ration $g_{2p}/gts = 3$. Thus,

$$f_{2s} = \frac{g_{2s}}{g_{2s}+g_{2p}} = \frac{1}{4} = \frac{x_{2s}}{x_{2p}+x_{2s}} = \frac{x_{2s}}{x_2}, \tag{4.67}$$

where $x_2 = x_{2s} + x_{2p}$ is the total population of level $n = 2$. Therefore, the evolution equations for x_{2p} and x_{2s} become

$$\begin{aligned}\frac{{\rm d}x_{2s}}{{\rm d}t} &= \frac{1}{4}\alpha_B {x_{\rm e}}^2 n_{\rm H+p} - \beta_B x_{2s} - A_{\rm 2s1s}\left(x_{2s} - x_{\rm 1s}e^{-h\nu_\alpha/(k_{\rm B}T)}\right)\\ \frac{{\rm d}x_{2p}}{{\rm d}t} &= \frac{3}{4}\alpha_B {x_{\rm e}}^2 n_{\rm H+p} - \beta_B x_{2p} - P_{\rm esc}A_{\rm 2p1s}\left[x_{2p} - 3x_{\rm 1s}e^{-h\nu_\alpha/(k_{\rm B}T)}\right].\end{aligned} \tag{4.68}$$

We can add the two equations to get the evolution equation for $x_2 = x_{2s} + x_{2p}$,

$$\begin{aligned}\frac{{\rm d}x_2}{{\rm d}t} &= \alpha_B {x_{\rm e}}^2 n_{\rm H+p} - \beta_B x_2 - A_{\rm 2s1s}\left(\frac{x_2}{4} - x_{\rm 1s}e^{-h\nu_\alpha/(k_{\rm B}T)}\right) - 3P_{\rm esc}A_{\rm 2p1s}\left[\frac{x_2}{4} - x_{\rm 1s}e^{-h\nu_\alpha/(k_{\rm B}T)}\right]\\ &= \alpha_B {x_{\rm e}}^2 n_{\rm H+p} + \left(A_{\rm 2s1s} + 3P_{\rm esc}A_{\rm 2p1s}\right)x_{\rm 1s}e^{-h\nu_\alpha/(k_{\rm B}T)} - x_2\left(\beta_B + \frac{1}{4}A_{\rm 2s1s} + \frac{3}{4}P_{\rm esc}A_{\rm 2p1s}\right)\end{aligned} \tag{4.69}$$

The transitions from $n = 2$ to $n = 1$ are fast compared to the recombination rate and the expansion rate. We can therefore assume that a steady state is reached for the $n = 2$ level, i.e. the recombinations to the $n = 2$ level are immediately balanced by the transitions to $n = 1$ level and the net change in the $n = 2$ level population is negligible, $\mathrm{d}x_2/\mathrm{d}t \approx 0$. In other words, the fast de-excitations of $n = 2$ level are able to balance any rate of recombination which populate these levels reaching a steady state very quickly compared to the expansion rate. Note that the outgoing rate from any level i cannot be larger than the incoming rate since the level population $x_i \geq 0$. For the second level, we will expect a tiny non-zero population corresponding to the electrons which are in transit through the level. In the steady state approximation, we easily obtain the solution to Eq. 4.69 as

$$x_2 = \frac{\alpha_B {x_\mathrm{e}}^2 n_{\mathrm{H+p}} + \left(A_{\mathrm{2s1s}} + 3P_{\mathrm{esc}}A_{\mathrm{2p1s}}\right) x_{\mathrm{1s}} e^{-h\nu_\alpha/(k_\mathrm{B}T)}}{\beta_B + \frac{1}{4}A_{\mathrm{2s1s}} + \frac{3}{4}P_{\mathrm{esc}}A_{\mathrm{2p1s}}} \tag{4.70}$$

The evolution equation for the $1s$ level is

$$\frac{\mathrm{d}x_{\mathrm{1s}}}{\mathrm{d}t} = A_{\mathrm{2s1s}}\left(x_{2s} - x_{\mathrm{1s}}e^{-h\nu_\alpha/(k_\mathrm{B}T)}\right) + P_{\mathrm{esc}}A_{\mathrm{2p1s}}\left(x_{2p} - 3x_{\mathrm{1s}}e^{-h\nu_\alpha/(k_\mathrm{B}T)}\right), \tag{4.71}$$

where we have taken into account the excitations and de-excitations from the $n = 2$ level and ignored direct recombinations and ionizations from the $1s$ level. The change is the free electron density is just the negative of the change in the $1s$ level, $\mathrm{d}x_\mathrm{e}/\mathrm{d}t = -\mathrm{d}x_{\mathrm{1s}}/\mathrm{d}t$, and using the solution we obtained above for $x_2 = 4x_{2s} = (4/3)x_{2p}$, we get

$$\begin{aligned}
\frac{\mathrm{d}x_\mathrm{e}}{\mathrm{d}t} &= -\frac{x_2}{4}\left(A_{\mathrm{2s1s}} + 3P_{\mathrm{esc}}A_{\mathrm{2p1s}}\right) + x_{\mathrm{1s}}e^{-h\nu_\alpha/(k_\mathrm{B}T)}\left(A_{\mathrm{2s1s}} + 3P_{\mathrm{esc}}A_{\mathrm{2p1s}}\right)\\
&= -\frac{\frac{1}{4}A_{\mathrm{2s1s}} + \frac{3}{4}P_{\mathrm{esc}}A_{\mathrm{2p1s}}}{\beta_B + \frac{1}{4}A_{\mathrm{2s1s}} + \frac{3}{4}P_{\mathrm{esc}}A_{\mathrm{2p1s}}}\left[\alpha_B {x_\mathrm{e}}^2 n_{\mathrm{H+p}} - 4x_{\mathrm{1s}}\beta_B e^{-h\nu_\alpha/(k_\mathrm{B}T)}\right]\\
&= -C\alpha_B\left[{x_\mathrm{e}}^2 n_{\mathrm{H+p}} - x_{\mathrm{1s}}\left(\frac{m_\mathrm{e}k_\mathrm{B}T}{2\pi\hbar^2}\right)^{3/2} e^{-E_{\mathrm{I},2}/(k_\mathrm{B}T)}e^{-h\nu_\alpha/(k_\mathrm{B}T)}\right]\\
&= -C\alpha_B\left[{x_\mathrm{e}}^2 n_{\mathrm{H+p}} - x_{\mathrm{1s}}\left(\frac{m_\mathrm{e}k_\mathrm{B}T}{2\pi\hbar^2}\right)^{3/2} e^{-E_\mathrm{I}/(k_\mathrm{B}T)}\right]\\
C &\equiv \frac{\frac{1}{4}A_{\mathrm{2s1s}} + \frac{3}{4}P_{\mathrm{esc}}A_{\mathrm{2p1s}}}{\beta_B + \frac{1}{4}A_{\mathrm{2s1s}} + \frac{3}{4}P_{\mathrm{esc}}A_{\mathrm{2p1s}}}.
\end{aligned} \tag{4.72}$$

Comparing it with the case B recombination, Eq. 4.41, we see that the effective recombination is slowed down, with $C < 1$. The factor C is the probability that a electron in $n = 2$ shell will eventually de-excite to the ground state instead of being ionized, and takes into account that the Ly-α photons cannot escape and the slow $2s-1s$ transition is important. In equilibrium, we again recover the Saha solution. The departure from equilbrium will occur even earlier compared to the case B recombination. The ODE describing the evolution of the electron fraction can be solved numerically once we know the Ly-α escape probability, $P_{\rm esc}$.

Ly-α escape probability

To find $P_{\rm esc}$, we must solve the radiative transfer equation,

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}t} = -\alpha_\nu I_\nu = -n_{1s}\sigma_{1s2p}cI_\nu. \tag{4.73}$$

The Lyman-α photons redshift with time due to the expansion of the Universe. We want to follow a photon emitted at a particular time until it is absorbed or escapes to infinity. For such a photon, there is a monotonic relationship between its energy with, $\nu \propto a(t)$, and time. We can thus use its frequency as a time variable and write

$$\begin{aligned}
\frac{\mathrm{d}I_\nu}{\mathrm{d}t} &= \frac{\mathrm{d}I_\nu}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}t} = \frac{\mathrm{d}I_\nu}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}a}\frac{\mathrm{d}a}{\mathrm{d}t} \\
&= \frac{\mathrm{d}I_\nu}{\mathrm{d}\nu}(-\nu H) = -n_{1s}\sigma_{1s2p}cI_\nu \\
&= -n_{1s}\frac{3c^3}{8\pi\nu^2}A_{\rm 2p1s}I_\nu\phi(\nu) \\
\frac{\mathrm{d}I_\nu}{\mathrm{d}\nu} &= \frac{1}{H}n_{1s}\frac{3c^3}{8\pi\nu^3}A_{\rm 2p1s}I_\nu\phi(\nu).
\end{aligned} \tag{4.74}$$

The solution to the above equation, with initial frequency of the photon ν_i is given by

$$\ln\left(\frac{I_\nu}{I_{\nu_i}}\right) = \frac{3c^3A_{\rm 2p1s}}{8\pi}\int_{\nu_i}^{\nu}\frac{\phi(\nu')}{H\nu'^3}\mathrm{d}\nu', \tag{4.75}$$

where the evolution is along $\nu \propto a^{-1}$, therefore H is also a function of ν. Alternatively, we can look at intensity as a function of two variable, (ν, t),$\mathrm{d}/\mathrm{d}t$ is the total derivative and the radiative transfer equation is a partial differential equation with two independent variables. The $\nu \propto a^{-1}$ represent the characteristic curves of the PDE along which the PDE reduces to an ODE and the above solution is the solution along the characteristic curve with initial date ν_i at initial time t_i or scalar factor a_i.

The profile $\phi(\nu)$ is usually very narrow with typical Doppler width defined by the thermal velocity $\upsilon_{\rm th}$ is given by (Eq. 3.16)

$$\frac{\Delta\nu_{\rm D}}{\nu_0} = \sqrt{\frac{2k_{\rm B}T}{m_{\rm H}c^2}} \sim 10^{-3}. \tag{4.76}$$

The natural line width is

$$\frac{\Delta\nu}{\nu} = \frac{\gamma}{2\pi\nu_0} \approx 4\times 10^{-8} \tag{4.77}$$

Exercise 18

Show that the Full Width at Half Maximum (FWHM) for the Lorentz profile, Eq. 3.14, is given by

$$\left.\frac{\Delta\nu}{\nu}\right|_{\rm FWHM} = \frac{\gamma}{2\pi\nu_0} \tag{4.78}$$

We can therefore assume that $\nu \approx \nu_0$ and $H \approx$ constant over the line profile in $H\nu^3$ factor in the integrand in Eq. 4.75. The contribution to the integral vanishes rapidly away from the line center because of $\phi(\nu)$. The solution can now be written as

$$I_\nu = I_{\nu_i} e^{-\tau_{\rm S}\int_\nu^{\nu_i}\phi(\nu'){\rm d}\nu'}, \tag{4.79}$$

where we have defined the Sobolev optical depth

$$\tau_{\rm S} = \frac{3c^3 A_{\rm 2p1s} n_{1s}}{8\pi H\nu_0^3}. \tag{4.80}$$

This escape of line photons in an expanding medium or a medium with a velocity gradient was studied by Sobolev for the expanding atmosphere of stars [38]. The solution is however very general and applies to any medium with a velocity gradient when considering line emission. The escape probability for a photon with an initial frequency ν_i can be read off from the above solution by letting the photon escape to infinity with $\nu \to 0$ as

$$P_{\rm esc}(\nu_i) = e^{-\tau_{\rm S}\int_0^{\nu_i}\phi(\nu){\rm d}\nu}. \tag{4.81}$$

In reality the photon will have escaped once it is sufficiently far from the line center so that $\phi(\nu)$ is negligible. We have extended the limit of integration to $\nu = 0$, since once *outside* the line, the absorption probability ($\propto \phi(\nu)$) and therefore integrand vanishes anyway. The initial frequency of the photon, ν_i is also distributed according to the line profile, $\phi(\nu_i)$. Averaging over all initial frequencies therefore gives the average escape probability,

$$\begin{aligned}\langle P_{\rm esc}\rangle &= \int_0^\infty P_{\rm esc}(\nu_i)\phi(\nu_i){\rm d}\nu_i\\ &= \int_0^\infty {\rm d}\nu_i\phi(\nu_i)e^{-\tau_{\rm S}\int_0^{\nu_i}\phi(\nu){\rm d}\nu}\end{aligned} \tag{4.82}$$

We can evaluate the integral by doing a change of variables,

$$\begin{aligned}x &= \int_0^{\nu_i}\phi(\nu){\rm d}\nu,\\ {\rm d}x &= \phi(\nu_i){\rm d}\nu_i\end{aligned} \tag{4.83}$$

the formula for average $P_{\rm esc}$ becomes

$$\langle P_{\rm esc}\rangle = \int_0^1 {\rm d}x e^{-\tau_{\rm S}x} = \frac{1-e^{-\tau_{\rm S}}}{\tau_{\rm S}}. \tag{4.84}$$

This is also known as the Sobolev escape probability.

Exercise 19

Calculate the factor C and escape probability $P_{\rm esc}$ at the freezeout redshift $z_{\rm f} = 1160$. This freezeout redshift was derived for case B recombination. What is the new freezeout redshift for 3-level atom ? Using the new freezeout redshift, calculate the new resdiual electron fraction for the recombination with 3-level atom model.

It turns out that when we do the full calculation, resolving all atomic levels upto a large n, and solve for all level populations explictly, there is a small speedup in recombination w.r.t the 3-level atom calculation [39]. Recent more precise calculation take into account many additional effects from atomic physics and radiative transfer to achieve a sub-percent precision [34, 40] and are accessible using publicly available codes CosmoRec and HyRec.

Each atomic transition, as the electron cascades towards the ground state after initially recombining to an excited state, results in emission of a recombination line photon. These photons, including the Lyman series photons which escape, form the cosmological recombination spectrum [41]. The shape of the cosmological recombination spectrum, just like the CMB anisotropy power spectrum, is sensitive to the cosmological parameters such as baryon density and helium fraction. The extraction of this cosmological information will be extremely challenging, since the additional photons per CMB blackbody photon that are produced during the cosmological recombination ($\sim$ 5 photons per hydrogen atom [35]) are of order baryon to photon ratio, $\eta \sim 10^{-9}$, resulting in a distortion from the blackbody spectrum of order nK. This can be compared to the 100 μK CMB fluctuations, which we can measure today with % level precision.

Chapter 5

Newtonian hydrodynamics

Continuity equation

Lets consider a fliud with density $\rho(\mathbf{r}, t)$ and velocity $\boldsymbol{v}(\mathbf{r}, t)$. The mass of the fluid in some volume element V_0 is $\int_{V_0} \rho \mathrm{d}V$. The mass flux through a surface element $\mathrm{d}\mathbf{S}$ is $\rho\boldsymbol{v}.\mathrm{d}\mathbf{S}$, where $\mathrm{d}\mathbf{S}$ points outward from the enclosed volume. The change in mass in the volume V_0 is

$$-\oint \rho\boldsymbol{v} \cdot \mathrm{d}\mathbf{S} = \frac{\partial}{\partial t}\int \rho \mathrm{d}V, \tag{5.1}$$

where the surface integral is over the closed surface enclosing volume V_0 and volume integral is over the volume V_0. Converting the surface integral to volume integral using the divergence theorem gives

$$\int \left(\frac{\partial \rho}{\partial t} + \nabla \cdot \rho\boldsymbol{v}\right) \mathrm{d}V = 0. \tag{5.2}$$

Since the above equality holds for any volume, the integrand must vanish giving us the continuity equation or equation of mass conservation,

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot \rho\boldsymbol{v} = 0 \\ \frac{\partial \rho}{\partial t} + \rho \nabla \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \nabla \rho = 0 \end{aligned} \tag{5.3}$$

Euler's equation

The force acting on a surface element $\mathrm{d}\mathbf{S}$ of fluid is $-P\mathrm{d}\mathbf{S}$, where P is the pressure and minus sign means that the force acting inward is positive. The net force on the volume element is given by intergrating the force over the full surface enclosing the volume,

$$-\oint P\mathrm{d}\mathbf{S} = -\int \nabla P \mathrm{d}V, \tag{5.4}$$

where we again used divergence theorem. The force per unit volume is therefore just $-\nabla P$. The equation of motion for an infinitesimal volume element is given by the Newton's second law,

$$\rho\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = -\nabla P \tag{5.5}$$

The derivative in the above equation, $\mathrm{d}/\mathrm{d}t$ is a total derivative and acts on implicit as well as explicit dependence on time. The Eq. 5.5 gives the force on a fluid element or particle as it moves in space. It is also called comoving derivative. Since velocity is a function of position as well as time, $v(\mathbf{r}, t)$,

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} &= \left.\frac{\partial \boldsymbol{v}}{\partial t}\right|_{\mathbf{r}} + \left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\cdot\nabla\right)\boldsymbol{v} \\ &= \left.\frac{\partial \boldsymbol{v}}{\partial t}\right|_{\mathbf{r}} + \boldsymbol{v}\cdot\nabla\boldsymbol{v}\end{aligned} \tag{5.6}$$

The above relation is just the chain rule, which in cartesian coordinates (x, y, z) is

$$\frac{\mathrm{d}}{\mathrm{d}t} \equiv \frac{\partial}{\partial t} + \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\partial}{\partial x} + \frac{\mathrm{d}y}{\mathrm{d}t}\frac{\partial}{\partial y} + \frac{\mathrm{d}z}{\mathrm{d}t}\frac{\partial}{\partial z} = \frac{\partial}{\partial t} + \boldsymbol{v}\cdot\nabla \tag{5.7}$$

The first term gives the change in velocity at same position and the second term gives the change in velocity because a fluid element moves to a space coordinate with a different velocity.

We want to study self gravitating systems and must include the gravitational force,

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v}\cdot\nabla\boldsymbol{v} = -\frac{1}{\rho}\nabla P - \nabla\phi \tag{5.8}$$

The gravitational potential, ϕ, is given by the Poisson's equation

$$\nabla^2\phi = 4\pi G\rho \tag{5.9}$$

We now have the set of three PDEs, Eqs. 5.3, 5.8, and 5.9 for four functions to be determined, $\rho, P, \boldsymbol{v}$, and ϕ. We must specify the equation of state of the fluid, $w = \rho/P$ to close the system of equations. More generally, we should specify pressure as a function of density, $P(\rho)$.

We are interested in studying the gravitational instability and find conditions under which small perturbations can grow and collapse under their self-gravity to form bound objects such as dark matter halos. Let us consider small perturbations around an average or zeroth order solution (denoted by subscript 0),

$$\begin{aligned}
\rho(\mathbf{r},t) &= \rho_0(\mathbf{r},t) + \delta\rho(\mathbf{r},t)\\
P(\mathbf{r},t) &= P_0(\mathbf{r},t) + \delta P(\mathbf{r},t)\\
\boldsymbol{v}(\mathbf{r},t) &= \boldsymbol{v}_0(\mathbf{r},t) + \delta\boldsymbol{v}(\mathbf{r},t)\\
\phi(\mathbf{r},t) &= \phi_0(\mathbf{r},t) + \delta\phi(\mathbf{r},t),
\end{aligned} \tag{5.10}$$

where average quantities are only a function of time but perturbations are functions of spatial position also. Substituting in Eqs. 5.3, 5.8, and 5.9 and keeping only linear order terms we get the hydrodynamic equations for first order perturbations,

$$\begin{aligned}
\frac{\partial\delta\rho}{\partial t} + \nabla\cdot(\rho_0\delta\boldsymbol{v}) + \nabla\cdot(\delta\rho\boldsymbol{v}_0) &= 0\\
\frac{\partial\delta\boldsymbol{v}}{\partial t} + (\boldsymbol{v}_0\cdot\nabla)\,\delta\boldsymbol{v} + (\delta\boldsymbol{v}\cdot\nabla)\,\boldsymbol{v}_0 &= \frac{\delta\rho}{\rho_0^2}\nabla P_0 - \frac{1}{\rho_0}\nabla\delta P - \nabla\delta\phi\\
\nabla^2\delta\phi &= 4\pi G\delta\rho\\
\delta P &= \frac{\mathrm{d}P}{\mathrm{d}\rho}\delta\rho \equiv {c_{\mathrm{s}}}^2\delta\rho,
\end{aligned} \tag{5.11}$$

where ${c_{\mathrm{s}}}^2 = \mathrm{d}P/\mathrm{d}\rho$ is the sound speed. We are dealing with an ideal fluid and therefore do not have any friction or dissipation terms.

The zeroth order variables must of course satisfy the original hydrodynamics equation and therefore pure zeroth order terms cancel when we substitute the first order expansions in Eq. 5.10 in Eqs. 5.3, 5.8, and 5.9 which reduce to purely first order equations in Eqs. 5.11. The validity of Eqs. 5.11 therefore depends on the existence of zeroth order solutions which are being perturbed.

It is tempting to consider as a zeroth order solution a static infinite homogeneous state and look at behaviour of small perturbations. However, just as with Friedmann equations, a static infinite homogeneous state cannot exist in Newtonian world either. In particular such a state will not satisfy the zeroth order Newtonian hydrodynamic equations. For $P_0 =$ constant, $\rho_0 =$ constant, $\boldsymbol{v}_0 = 0$, Eq. 5.8 gives $\nabla\phi_0 = 0$ while Poisson's equation gives $\nabla^2\phi_0 = 4\pi G\rho_0$. However, in astrophysics we usually ignore this contradiction and study Eq. 5.11 for perturbations around a static homogeneous fluid. This inconsistency can be *argued* away by an assumption called *Jeans swindle* [42] after Jeans who first studied gravitational instabilities in an infinite medium [43]. The assumption is that Poisson's equation is applicable only to perturbed density and potential and the unperturbed potential is zero.

We could also argue that a correct description of an infinite medium needs a relativistic theory and therefore the arbitrary setting of unperturbed potential to zero in Newtonian description is justified. In particular, we should really be perturbing around the correct zeroth order solution in general relativity, which are the Friedmann solutions. Nevertheless, we still get useful insights into gravitational collapse in Newtonian approximation. The results agree with a relativistic calculation where we have a consistent zeroth order solution. With the zeroth order state set to ρ_0 = constant, P_0 = constant, $\boldsymbol{v}_0 = 0, \phi_0 = 0$, the first order equations become (with $\boldsymbol{v} = \delta\boldsymbol{v}, \phi = \delta\phi$),

$$\frac{\partial\delta\rho}{\partial t} + \rho_0 \nabla\cdot\boldsymbol{v} = 0 \tag{5.12}$$

$$\begin{aligned}\frac{\partial\boldsymbol{v}}{\partial t} &= -\frac{1}{\rho_0}\nabla\delta P - \nabla\phi \\ &= -\frac{{c_s}^2}{\rho_0}\nabla\delta\rho - \nabla\phi \end{aligned} \tag{5.13}$$

$$\nabla^2\phi = 4\pi G\delta\rho \tag{5.14}$$

Taking time derivative of Eq. 5.12 and gradient of Eq. 5.13, we can eliminate velocity term to get a second order PDE for density,

$$\frac{\partial^2\delta\rho}{\partial t^2} - {c_s}^2\nabla^2\delta\rho - 4\pi G\rho_0\delta\rho = 0 \tag{5.15}$$

Without gravity, we just have the usual wave equation yielding sound waves as the solution. We can solve Eq. 5.15 by doing a Fourier transform or substituting the plane wave solutions, $\delta\rho = \rho_{\mathbf{k}} e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$ yiedling the dispersion relation

$$\begin{aligned}-\omega^2 + {c_s}^2 k^2 - 4\pi G\rho_0 &= 0 \\ \omega &= \left({c_s}^2 k^2 - 4\pi G\rho_0\right)^{1/2}\end{aligned} \tag{5.16}$$

The time evolution of any plane wave perturbation with spatial wavenumber $\mathbf{k}$ is determined by the frequency ω. We will have sound waves if ω is real and exponential growth, $\delta\rho \propto e^{2\pi t/\tau}$, if $\omega = -2\pi i/\tau$ is imaginary. The boundary between these two solution is given by vanishing of ω at Jeans wavelength, $\lambda_J = 2\pi/k_J$,

$$k_J = \sqrt{\frac{4\pi G\rho_0}{{c_s}^2}} \tag{5.17}$$

The mass inside a sphere of diameter λ_J is called Jeans mass and it is the smallest mass object in which gravity can overcome thermal pressure resulting in gravitational collapse,

$$M_J = \frac{4\pi}{3}\left(\frac{\lambda_J}{2}\right)^3 \rho_0 = \frac{\rho_0 \pi}{6}\left(\frac{\pi {c_s}^2}{G\rho_0}\right)^{3/2} \tag{5.18}$$

We will see that most of the above results survive in an expanding Universe. The most important change is that on scales unstable to gravitational collapse, i.e. scales larger than the Jeans length, the growth once the collapse starts is not exponential in an expanding or contracting Universe but a power law. This was first pointed out by Lifshitz in his seminal paper in 1946 [44].

We can include the effect of expansion as follows [45]. Note that in the static infinite Universe used by Jeans for the zeroth order solution, all zeroth order quantities are also independent of time. In an expanding Universe, the zeroth order quantity, such as density would be functions of time, $\rho_0(t)$ etc. Lets change the spatial coordinates from physical coordinates $\mathbf{r}$ to comoving coordinates $\mathbf{x}$ which do not depend on time for comoving observers,

$$\begin{aligned} \mathbf{r}(t) &= a(t)\mathbf{x} \\ \boldsymbol{v} &= \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \dot{a}\mathbf{x} + a\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \boldsymbol{v}_0 + \delta\boldsymbol{v}, \end{aligned} \tag{5.19}$$

where $\boldsymbol{v}_0$ is the veolocity due to Hubble expansion and $\delta\boldsymbol{v}$ is the peculiar velocity. We also need to transform the partial derivatives. The spatial derivatives are easy, denoting $\nabla_{\mathbf{r}}$ as the derivative operator in $\mathbf{r}$ coordinates and $\nabla_{\mathbf{x}}$ in $\mathbf{x}$ coordinates, since the scale factor $a(t)$ does not depend on the spatial coordinates,

$$\nabla_{\mathbf{r}} = \frac{1}{a}\nabla_{\mathbf{x}}. \tag{5.20}$$

The total time derivative is given by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} &= \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}} + \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\cdot\nabla_{\mathbf{x}} \\ &= \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}} + \frac{\delta\boldsymbol{v}}{a}\cdot\nabla_{\mathbf{x}} \end{aligned} \tag{5.21}$$

At zeroth order, total time derivative is just the partial time derivaive evaluated at constant values of comoving coordinates $\mathbf{x}$. The partial time derivative at constant $\mathbf{r}$ is also related to the partial time derivative at constant $\mathbf{x}$ using the chain rule, for any function $f(\mathbf{r}(\mathbf{x},t),t)$,

$$\begin{aligned}\left.\frac{\partial f}{\partial t}\right|_{\mathbf{x}} &= \left.\frac{\partial f}{\partial t}\right|_{\mathbf{r}} + \left.\frac{\partial \mathbf{r}}{\partial t}\right|_{\mathbf{x}} \cdot \nabla_{\mathbf{r}} f \\ &= \left.\frac{\partial f}{\partial t}\right|_{\mathbf{r}} + \dot{a}\mathbf{x} \cdot \nabla_{\mathbf{r}} f \\ \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}} &= \left.\frac{\partial}{\partial t}\right|_{\mathbf{r}} + \frac{\dot{a}}{a}\mathbf{x} \cdot \nabla_{\mathbf{x}} \\ &= \left.\frac{\partial}{\partial t}\right|_{\mathbf{r}} + \frac{\boldsymbol{v}_0}{a} \cdot \nabla_{\mathbf{x}} \end{aligned} \tag{5.22}$$

We note that for a homogeneous fluid, the gradients vanish and the two partial time derivatives coincide.

We can now write down the hydrodynamic equations at zeroth order. The continuity equation is

$$\frac{\partial \rho_0}{\partial t} + \frac{\rho_0}{a}\nabla_{\mathbf{x}} \cdot \boldsymbol{v}_0 = 0, \tag{5.23}$$

where $\nabla_{\mathbf{x}} \cdot \boldsymbol{v}_0 = \dot{a}\nabla_{\mathbf{x}} \cdot \mathbf{x} = 3\dot{a}$, giving

$$\begin{aligned}\frac{\partial \rho_0}{\partial t} + 3\frac{\dot{a}}{a}\rho_0 &= 0 \\ \rho_0 &\propto a^{-3}\end{aligned} \tag{5.24}$$

The Euler's equation at zeroth order is (see Eq. 5.21)

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{v}_0}{\mathrm{d}t} &= \left.\frac{\partial \boldsymbol{v}_0}{\partial t}\right|_{\mathbf{x}} \\ &= -\frac{1}{\rho_0}\frac{1}{a}\nabla_{\mathbf{x}}P - \frac{1}{a}\nabla_{\mathbf{x}}\phi_0 \\ \Rightarrow \mathbf{x}\ddot{a} &= -\frac{1}{a}\nabla_{\mathbf{x}}\phi_0\end{aligned} \tag{5.25}$$

The Poisson equation at zeroth order is

$$\frac{1}{a^2}\nabla^2_{\mathbf{x}}\phi_0 = 4\pi G\rho - \Lambda, \tag{5.26}$$

where we have included an extra constant which corresponds to the cosmological constant in general relativity. This is the Newtonian limit of the Einstein's equations with a cosmological constant. Taking divergence of Eq. 5.25 we get,

$$\frac{1}{a}\nabla_{\mathbf{x}} \cdot \mathbf{x}\ddot{a} = -\frac{1}{a^2}\nabla^2_{\mathbf{x}}\phi_0 = -4\pi G\rho + \Lambda \tag{5.27}$$

We thus recover the Friedmann deceleration equation without the pressure terms, since $\nabla_{\mathbf{x}} \cdot \mathbf{x} = 3$,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G\rho}{3} + \frac{\Lambda}{3} \tag{5.28}$$

The pressure term is still missing from the above equation, which is expected since we are working in the Newtonian limit. Note that for a honogeneous infinite Universe, the absolute value of ϕ_0 still has no meaning. However, we have a consistent relation between its derivatives and thus a consistent zeroth order solution, unlike the static Universe solution.

At first order we get for the continuity equation,

$$\left.\frac{\partial\delta\rho}{\partial t}\right|_{\mathbf{r}} + \frac{\rho_0}{a}\nabla_{\mathbf{x}} \cdot \delta\boldsymbol{v} + \frac{\delta\rho}{a}\nabla_{\mathbf{x}} \cdot \boldsymbol{v}_0 + \frac{\boldsymbol{v}_0}{a} \cdot \nabla_{\mathbf{x}}\delta\rho = 0 \tag{5.29}$$

Using equation 5.22 and $\nabla_{\mathbf{x}}.\boldsymbol{v}_0 = 3\dot{a}$, we get

$$\left.\frac{\partial\delta\rho}{\partial t}\right|_{\mathbf{x}} + \frac{\rho_0}{a}\nabla_{\mathbf{x}} \cdot \delta\boldsymbol{v} + 3\delta\rho\frac{\dot{a}}{a} = 0 \tag{5.30}$$

The Euler equation at first order is

$$\left.\frac{\partial\delta\boldsymbol{v}}{\partial t}\right|_{\mathbf{x}} + \left(\frac{\delta\boldsymbol{v}}{a} \cdot \nabla_{\mathbf{x}}\right)\mathbf{x}\dot{a} = -\frac{\nabla_{\mathbf{x}}\delta\rho {c_s}^2}{a\rho_0} - \frac{\nabla_{\mathbf{x}}}{a}\delta\phi$$
$$\left.\frac{\partial\delta\boldsymbol{v}}{\partial t}\right|_{\mathbf{x}} + \frac{\dot{a}}{a}\delta\boldsymbol{v} = -\frac{{c_s}^2}{a\rho_0}\nabla_{\mathbf{x}}\delta\rho - \frac{\nabla_{\mathbf{x}}}{a}\delta\phi \tag{5.31}$$

The Poisson equation at first order is

$$\nabla^2_{\mathbf{x}}\delta\phi = 4\pi G a^2\delta\rho \tag{5.32}$$

Taking the curl of Eq. 5.31, we see that the right hand side vanishes as it has pure gradients of scalars,

$$\left.\frac{\partial\nabla \times \delta\boldsymbol{v}}{\partial t}\right|_{\mathbf{x}} + \frac{\dot{a}}{a}\nabla \times \delta\boldsymbol{v} = 0$$
$$\nabla \times \delta\boldsymbol{v} \propto a^{-1} \tag{5.33}$$

The curl modes decay with expansion since they have no source. This is just a consequence of redshifting of momenta with expansion. The scalar modes are sourced by gravity (i.e. matter is continuously being accelerated by gravitational forces from near by matter) of matter perturbations and therefore survive. This also means, that if we decompose initial velocity field into a sum of a pure vector field with vanishing divergence (or vector modes or curl modes) and a velocity field with vanishing curl and which therefore can be written as a divergence of a scalar velocity potential (i.e. scalar mode of velocity), the curl mode will soon redshift away and the total velocities would be purely scalar modes. It is convenient to define fractional density perturbation, $\delta\rho = \delta \times \rho_0$, with $\rho_0 \propto a^{-3}$. The continuity equation for δ becomes

$$\left.\frac{\partial \delta}{\partial t}\right|_{\mathbf{x}} + \frac{\rho_0}{a}\nabla_{\mathbf{x}} \cdot \delta \boldsymbol{v} = 0 \tag{5.34}$$

As before, taking divergence of Eq. 5.31 (applying $\nabla_{\mathbf{x}}/a$) and time derivative of Eq. 5.34 and subtracting we get

$$\frac{\partial^2 \delta}{\partial t^2} - 2\frac{\dot{a}}{a^2}\nabla_{\mathbf{x}} \cdot \delta \boldsymbol{v} = 4\pi G \rho_0 \delta + \frac{{c_s}^2}{a^2}\nabla_{\mathbf{x}}^2 \delta, \tag{5.35}$$

where the partial time derivatives from now on are understood to be at constant $\mathbf{x}$ and we have dropped explicit labels specifying it from the equations. Using the continuity once again to eliminate the remaining velocity term we get,

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G \rho_0 \delta + \frac{{c_s}^2}{a^2}\nabla_{\mathbf{x}}^2 \delta \tag{5.36}$$

We therefore have an equation very similar to the one we obtained for the Jeans analysis in a static Universe, with one extra *Hubble friction* term $\propto H$. In other words, instead of the simple oscillltor equation earlier, we now have a damped oscillator with the expansion of the Universe providing the damping. We can do again analysis similar to the Jeans analysis by taking the Fourier transform or substituting the trial solution $\delta_{\mathbf{k}} = e^{i\omega t - \mathbf{k} \cdot \mathbf{x}}$. Note that now $\mathbf{k}$ is the Fourier counterpart of the comoving coordinates $\mathbf{x}$ and is therefore the comoving wavenumber. The physical wavenumber is given by $\mathbf{k}/a$. Substituting the plain wave solution, we get the following dispersion relation,

$$-\omega^2 + 2iH\omega = 4\pi G \rho_0 - \frac{{c_s}^2 k^2}{a^2} \tag{5.37}$$

The frequency ω should therefore have a real part as well as an imaginary part. Writing $\omega = \alpha + i\beta$, and equating the real and imaginary parts of the dispersion relation, The imaginary part gives us

$$\beta = H, \tag{5.38}$$

and therefore we have a damping factor of the form e^{-Ht}. The real part now gives us

$$\alpha^2 = \frac{{c_s}^2 k^2}{a^2} - 4\pi G \rho_0 - H^2 \tag{5.39}$$

The comoving particle horizon is given by $(aH)^{-1}$ and $c_{\rm s}/(aH)$ is the comoving sound horizon, i.e. the maximum distance a sound wave in the medium could have traveled. For α to remain real, we should have $k \geq aH/c_{\rm s}$, i.e. we see that even in Newtonian limit, we have oscillations on scales small compared to the sound horizon. Note that there might be other components in the Universe, in addition to and decoupled from the fluid under consideration, contributing to the Hubble expansion, $H^2 = 8\pi G(\rho_0 + \sum_i \rho_i)/3 \geq 8\pi G\rho_0/3$, where ρ_i is the energy density in ith component. The Jeans scale in an expanding Universe is given by as before by vanishing of α,

$$\lambda_{\rm J} = \frac{2\pi}{k_{\rm J}} = \sqrt{\frac{4\pi^2 {c_{\rm s}}^2}{4\pi G\rho_0 + H^2}}. \tag{5.40}$$

For $k > k_{\rm J}$, or scales smaller than the Jeans scale, α is real and we have oscillations. For oscillations, the scale λ must be below not only the original Jeans scale, but also below the sound horizon. On scales greater than the sound horizon (which is of the same order or smaller compared to the Jeans scale including expansion but smaller than the static Jeans scale), we can have gravitational growth. For these scales, ω is purely imaginary. However, the growth is no longer exponential. To find a solution, we must also take into account that ρ_0 and H are also a functions of time. We can find solution for the scales much larger compared to the Jeans scale, when we can neglect the pressure term, in an expanding universe by doing a change of variables from t to a,

$$\begin{aligned}\frac{\partial\delta}{\partial t} &= \frac{\partial\delta}{\partial a}\frac{{\rm d}a}{{\rm d}t} = aH\frac{\partial\delta}{\partial a}\\ \frac{\partial^2\delta}{\partial t^2} &= a^2H^2\frac{\partial^2\delta}{\partial a^2} + aH^2\frac{\partial\delta}{\partial a} + a^2H\frac{{\rm d}H}{{\rm d}a}\frac{\partial\delta}{\partial a}\end{aligned} \tag{5.41}$$

For pure matter dominated Universe, with $H^2 = 8\pi G\rho_0/3$ and $\rho_0 \propto a^{-3}$, we have

$$2H\frac{{\rm d}H}{{\rm d}a} = \frac{8\pi G}{3}\frac{{\rm d}\rho_0}{{\rm d}a} = \frac{8\pi G}{3}\left(\frac{-3\rho_0}{a}\right) = \frac{-3H^2}{a} \tag{5.42}$$

Using these relations in Eq. 5.36 we get

$$\frac{\partial^2\delta}{\partial a^2} + \left(\frac{3}{a} + \frac{1}{H}\frac{{\rm d}H}{{\rm d}a}\right)\frac{\partial\delta}{\partial a} = \frac{4\pi G\rho_0\delta}{a^2H^2} \tag{5.43}$$

$$\frac{\partial^2\delta}{\partial a^2} + \frac{3}{2a}\frac{\partial\delta}{\partial a} - \frac{3}{2}\frac{\delta}{a^2} = 0 \tag{5.44}$$

The solutions are $\delta \propto a, a^{-3/2}$. Since in matter dominated era, $a \propto t^{2/3}$, the growing mode grows as a power law.

If we had a background dominated by vaccum energy, $\rho_\Lambda \gg \rho_{\rm m} \equiv \rho_0$, e.g. dark energy in the current Universe or inflaton field during inflation, then $H^2 \approx 8\pi G\rho_\Lambda/3$=constant and the evolution of matter perturbations in such a background becomes,

$$\begin{aligned}\frac{\partial^2\delta}{\partial a^2} + \frac{3}{a}\frac{\partial\delta}{\partial a} &= \frac{4\pi G\rho_0\delta}{a^2H^2}\\ a^2\frac{\partial^2\delta}{\partial a^2} + 3a\frac{\partial\delta}{\partial a} &= \frac{3\rho_0\delta}{2\rho_\Lambda}\end{aligned} \tag{5.45}$$

Fro vaccum energy domination, $\rho_0/\rho_\Lambda \to 0$. The solutions in this limit are $\delta(a) =$ *constant* and $\delta \propto a^{-2} \propto e^{-2Ht}$. Thus the decaying mode quickly disappears while the growing mode is just constant. The matter density perturbations do not grow during inflation or dark energy domination.

Exercise 20

During radiation dominated era, $H^2 \approx 8\pi G\rho_{\rm r}/3$, where the radiation energy density $\rho_{\rm r} \propto a^{-4}$. Solve Eq. 5.43 during radiation dominated era and show that the leading or growing mode is $\propto \ln a/a_{\rm eq}$ when $a/a_{\rm eq} \ll 1$, where $a_{\rm eq}$ is the scale factor corresponding to matter-radiation equality.
Hint: Note that ρ_0 in the numerator on the right hand side of Eq. 5.43 is equal to the dark matter density, while Hubble parameter is dominated by the radiation energy density when $a/a_{\rm eq} \ll 1$.

Note that during the contracting phase of the Universe, the decaying mode becomes growing mode, and during matter domination, perturbations *grow* as $\propto a^{-3/2}$.

Chapter 6

Relativistic hydrodynamics

We want to study the evolution of Friedmann Universe with small perturbations. As before, we will expand around the zeroth order solution which is the homogeneous and isotropic expanding *flat* Friedmann Universe which we will denote by the average metric $\bar{g}_{\mu\nu}$. We will assume the background spatial geometry to be flat for simplicity. This is a good assumption as all cosmological observations indicate that the Universe is flat and constrain the curvature to be very small.

6.1 Scalar vector tensor decomposition of perturbations

We write the full metric, $g_{\mu\nu}$ as sum of the Friedmann metric, $\bar{g}_{\mu\nu}$, and a small perturbation, $h_{\mu\nu}$,

$$
\begin{aligned}
g_{\mu\nu} &= \bar{g}_{\mu\nu} + h_{\mu\nu} \\
\bar{g}_{00} = -1, \bar{g}_{i0} &= \bar{g}_{0i} = 0, \bar{g}_{ij} = a^2\delta_{ij}
\end{aligned}
\tag{6.1}
$$

Exercise 21

Show that the inverse metric is given by (upto first order)

$$
\begin{aligned}
g^{\mu\nu} &= \bar{g}^{\mu\nu} + h^{\mu\nu} \\
h^{\mu\nu} &= -\bar{g}^{\mu\alpha}\bar{g}^{\nu\beta}h_{\alpha\beta},
\end{aligned}
\tag{6.2}
$$

where $\bar{g}^{\mu\nu}$ is the inverse of the background Friedmann metric.

We will be using the notation where we denote partial derivative w.r.t coordinate α by ", α" , e.g.

$$f_{\mu\nu,\alpha} \equiv \frac{\partial f_{\mu\nu}}{\partial x^\alpha} \tag{6.3}$$

for any tensor $f_{\mu\nu}$. In particular, all indices after the "," mean partial derivatives w.r.t all of those coordinates.

In general each component of $h_{\mu\nu}$ can be non zero. Since the metric perturbation is also a symmteric tensor, like the full metric tensor, it has 10 indepdent components. It is useful to decompose the perturbations into scalar, vector and tensor components. At first order in perturbations, the scalar, vector and tensor perturbations are decoupled from each other. In particular they evolve independently and can therefore be studied separately. Lets look at different components of $h_{\mu\nu}$.

The time-time component,

$$h_{00} = A, \tag{6.4}$$

is obviously a scalar which we denote by a scalar function A. The time-space components,

$$h_{0i} = B_i \tag{6.5}$$

are 3-vectors. Any vector B_i can be decomposed into a divergence free vector part and a gradient of a scalar,

$$B_i = v_i + u_{,i}, \tag{6.6}$$

where u is a scalar and

$$v_i^{;i} = v^i_{,i} = \nabla \cdot \mathbf{v} = 0 \tag{6.7}$$

Taking divergence of Eq. 6.6

$$B_i^{;i} = u_{,i}{}^{i} \tag{6.8}$$

This is an elliptic equation written in more familiar form as

$$\nabla^2 u = \nabla \cdot \mathbf{B} \tag{6.9}$$

The statement that a general vector field B can be separated into a scalar field u and a divergence free vector field $\mathbf{v}$ relies on existence of the solutions for the above elliptic equation. This equation always has a solution as long as u falls of sufficiently rapidly with distance $R \to \infty$ or if there is a boundary i.e space-time is compact as in the case of a closed Universe. We will assume that the vector fields we will encounter satisfy these conditions. The part u of the vector field B_i is the scalar perturbation and v_i is the vector component of the metric perturbation B_i.

The purely spatial part h_{ij} can similarly be decomposed into a trace part and another scalar, a divergence free vector, and a divergenceless traceless tensor part,

$$h_{ij} = C\delta_{ij} + \frac{\partial^2 D}{\partial x^i \partial x^j} + \frac{\partial E_i}{\partial x_j} + \frac{\partial E_j}{\partial x_i} + F_{ij}, \tag{6.10}$$

where E_i is a divergenceless vector,

$$\frac{\partial E_i}{\partial x_i} \equiv \nabla \cdot \mathbf{E} = 0. \tag{6.11}$$

Also, F_{ij} is symmetric ($F_{ij} = F_{ji}$), divergenceless ($F_{ij,i} = 0$), and traceless ($F_{ij}\delta^{ij} = F^i_i = 0$). Taking the trace of Eq. 6.10, we see that

$$h_{ij}\delta^{ij} = h^i_i = 3C + \delta^{ij}\frac{\partial^2 D}{\partial x^i \partial x^j} = 3C + \nabla^2 D. \tag{6.12}$$

Taking divergence of Eq. 6.10,we get

$$\frac{\partial h_{ij}}{\partial x^i} = \frac{\partial C}{\partial x^i}\delta_{ij} + \frac{\partial^3 D}{\partial x^{i^2} \partial x^j} + \frac{\partial^2 E_j}{\partial x^{i^2}}. \tag{6.13}$$

Differentiating again w.r.t x^j,

$$\frac{\partial h_{ij}}{\partial x^i \partial x^j} = \nabla^2 C + \nabla^2 \left(\nabla^2 D\right). \tag{6.14}$$

Eliminating $\nabla^4 D$ from Eqs. 6.14 and 6.12 gives us an elliptic solution for C for which a unique solution exists as long as the tensor field falls off fast enough at infinity as discussed above. Similarly, we must solve the elliptic equation twice to get the solution for D. Knowing C and D, we can solve Eq. 6.13 for E_j and Eq. 6.10 then gives F_{ij}. We have therefore shown that we can decompose the perturbations of any general tensor field into scalars (including spatial derivatives of scalars), divergenceless vectors and divergenceless traceless tensors. The last, when we are decomposing the metric perturbations h_{ij}, are of course the gravitational waves. This decomposition is very useful. It can be shown that at linear order, the scalar, vector, and tensor perturbations evolve independently of each other as long as the evolution equations are *tensor equations* and at most second order in differential operators [46]. Our equations, derived from Einstein's equations, satisfy all these criteria and we can therefore study them independently.

Intuitively, we can see that the different types of perturbations must decouple at linear order as follows. A scalar equation must have all terms appearing as scalars. The only way a vector can appear is through a scalar product with another vector. Also at linear order, a first order vector term can appear as a product with only a zeroth order term. At linear order the isotropy and homogeneity means that there is no preferred direction or vector at zeroth order which can appear in a scalar product with the first order vector perturbation except the derivative operator, $\partial/\partial x^i$. If we take the Fourier transform of equations, we have equivalently the wavevector $\mathbf{k}$ along which the perturbation is varying. By definition, the divergenceless vector is orthogonal to $\partial/\partial x^i$ or $\mathbf{k}$ ($\mathbf{k}.\boldsymbol{v} = 0$, where $\boldsymbol{v}$ is a divergenceless vector) and therefore cannot appear in the scalar equation. Similarly all terms in the vector equation would be orthogonal to $\mathbf{k}$, while the only way a scalar perturbation can appear in a vector equation is as a product with the only available zeroth order vector, $\mathbf{k}$. Similarly, equations for the evolution of tensor (divergenceless traceless) perturbations are decoupled from both scalar and vector perturbations.

Since the only persisting source of perturbations in standard ΛCDM cosmology is the scalar gravitational potential, the scalar equations are of the most importance and we will study them in detail. In particular, the scalar perturbations grow and lead to the formation of galaxies and the large scale structure arising from the clustering of the galaxies.

The vector modes, e.g. the divergenceless part of velocities or vortical motions, decay as $\propto a^{-1}$ in the absence of any source. Therefore even if vector modes were present in the early Universe, they decay away quickly and soon become irrelevant in linear theory. In non-standard cosmologies or extension of ΛCDM cosmology, we may have sources for vector modes with important observational consequences, e.g. in CMB polarization anisotropies. One interesting source of vector perturbations is cosmic strings (and other topological defects). These are also called active sources since they are continuously *churning* the matter through which they move, creating new perturbations (scalar, vector, and tensor) throughout the history of the Universe. This is in contrast with standard ΛCDM cosmology where perturbations are created only in the initial inflationary epoch of the Universe. Tensor perturbations are of course just the gravitational waves and can be primordial in origin (e.g. from inflation) or created by active sources.

6.1.1 Conformal Newtonian gauge

The scalar vector tensor decomposition tells us that there are four scalar degrees of freedom in the metric perturbation h_{ij}, the time-time component A, velocity potential u, trace of the spatial part, C and D. All four of these scalar degrees of freedom are however not physical. In general relativity, we have the freedom to choose coordinates and two metrics $g_{\mu\nu}$ and $g'_{\mu\nu}$ related by a coordinate transform $x^\mu \to x^\mu + \epsilon^\mu(x^\mu)$ are equivalent and describe the same space-time. We can use the freedom of coordinate choice to set two of the scalar *gauge* degrees of freedom to zero so that only two *physical* degrees of freedom remain. This called a gauge choice and there are many possibilties. One popular choice is the *conformal Newtonian gauge*, which sets u and D to zero. Another popular choice is called *synchronous gauge* in which the time components of the metric are unterturbed, $u = A = 0$. We will use conformal Newtonian gauge, as the evolution equations in this gauge resemble very closely to the equations of Newtonian hydrodynamics. The metric in conformal Newtonian gauge is give by

$$\mathrm{d}s^2 = -(1+2\psi)\,\mathrm{d}t^2 + a^2\,(1+2\phi)\,\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{6.15}$$

The signs of ϕ and ψ and are a matter of convention, we will use the same convention as Dodelson [47]. Some authors also interchange ϕ and ψ, i.e. the time-time component may be referred to as ϕ in some literature. We want to study the hydrodynamics in the presence of metric perturbations. The Poisson equation should of course be replaced by linearized Einstein's equation. The continuity and Euler equations however also get modified due to the presence of metric perturbations. We can derive these modified equations from conservation of stress-energy tensor. We will however take a different approach and derive them from Boltzmann equation which is more general and is valid even when the *fluid approximation* breaks down, as is the case for free streaming photons and neutrinos.

6.2 The Boltzmann equation

A rigorous derivation of the Boltzmann equation for photons can be found in [48]. In quantum theory the evolution of an operator O is given in the Heisenberg picture as

$$i\frac{\mathrm{d}O(t)}{\mathrm{d}t} = [O(t), H]\,, \tag{6.16}$$

where H is the Hamiltonian of the system and square brackets represent the commutator. In particular, we are interested in the evolution of the number operator,

$$N(p) = a^\dagger(p)a(p), \tag{6.17}$$

which counts the number of particles with momentum p, where $a^\dagger$ and a are the creation and annihilation operators. Its evolution is therefore given by

$$i\frac{\mathrm{d}N}{\mathrm{d}t} = [N, H] \tag{6.18}$$

We will deal with systems with large number of particles and would be interested in the expectation value of N, the occupation number of each state, $f \equiv \langle N \rangle$.

For a free field, a system of non-interacting particles, $[N, H] = 0$ and we recover Liouville's equation,

$$\frac{\mathrm{d}f}{\mathrm{d}t} = 0. \tag{6.19}$$

If there are interactions, they can result in change in the occupation numbers due to change in momenta of particles and also creation and destruction of particles, and we have the Boltzmann equation,

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}t} &= C(f)\\ iC(f) &\equiv \langle [N, H] \rangle\end{aligned} \tag{6.20}$$

The collision term $C(f)$ must be calculated from quantum field theory.

We want to derive the Boltzmann equations for each of the components of the Universe. If there are interactions between two particle species, their Boltzmann equations would be coupled. The Boltzmann equations are more general compared to the hydrodynamics equations we have considered so far. The energy and momentum conservation equations would follow from integrating the Boltzmann equations over momentum space.

In curved space-time, we must use affine parameter, λ, instead of the time coordinate t, giving

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}\lambda} &= C'(f)\\ C'(f) &= \frac{\mathrm{d}t}{\mathrm{d}\lambda}C(f)\end{aligned} \tag{6.21}$$

6.2.1 Liouville's operator in curved space time

We want to evaluate the Boltzmann equation and in particular the Liouville's operator, $\mathrm{d}/\mathrm{d}\lambda$ in curved space time. All information about curvature goes into the left hand side of the Boltzmann equation while the right hand side or the collision term contains all the particle physics. We will first evaluate

$$\frac{\mathrm{d}f}{\mathrm{d}\lambda}, \tag{6.22}$$

where f is the occupation number of the particle we are evolving. In general, we will have many particles with interactions among them and we will have a coupled system of Boltzmann equations, one for each particle, coupled by the collision terms. The occupation number is a function of spacetime coordinates and 3-momentum $\boldsymbol{p}$, $f \equiv f(t, \mathbf{x}, \boldsymbol{p})$, $\boldsymbol{p} = p\hat{\mathbf{p}}$. Expanding the total derivative we get

$$\frac{\mathrm{d}f}{\mathrm{d}\lambda} = \frac{\partial f}{\partial t}\frac{\mathrm{d}t}{\mathrm{d}\lambda} + \frac{\partial f}{\partial x^i}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda} + \frac{\partial f}{\partial p}\frac{\mathrm{d}p}{\mathrm{d}\lambda} + \frac{\partial f}{\partial \hat{p}^i}\frac{\mathrm{d}\hat{p}^i}{\mathrm{d}\lambda} \tag{6.23}$$

We want to evaluate the Liouville's operator at first order in perturbations using the perturbed metric, Eq. 6.15. The zeroth component of the 4-momentum is by definition $P^0 = \mathrm{d}t/\mathrm{d}\lambda$. We want to write this in terms of the energy E measured by a comoving observer with velocity V^μ in the perturbed Friedmann metric. The 4-velocity of a comoving observer should satisfy $g_{\mu\nu}V^\mu V^\nu = -1$. By definition the three component of the velocity of a comoving observer is equal to zero, $V^i = 0$. Therefore

$$\begin{aligned} -(1+2\psi){V^0}^2 &= -1 \\ V^0 = \frac{1}{\sqrt{1+2\psi}} &= 1-\psi, \end{aligned} \tag{6.24}$$

where we are doing Taylor series expansion in ψ and ϕ and only keeping terms upto first order in ψ, ϕ. The energy measured by a comoving observer is

$$\begin{aligned} E &= -g_{\mu\nu}P^\mu V^\nu = -g_{00}P^0V^0 \\ &= (1+2\psi)\,P^0\,(1-\psi) = (1+\psi)\,P^0 \\ P^0 &= E\,(1-\psi)\,. \end{aligned} \tag{6.25}$$

The energy measured by a comoving observer is therefore not the same as the zeroth component of 4-momentum which would be the energy measured in unperturbed Friedmann metric. The extra term ψ represents gravitational redshift. Similarly, the space component of the 4-momentum $P^i = \mathrm{d}x^i/\mathrm{d}\lambda$ can be evaluated as

$$\begin{aligned} g_{\mu\nu}P^\mu P^\nu &= -m^2 \\ -(1+2\psi)\,{P^0}^2 + a^2\,(1+2\phi)\sum_i {P^i}^2 &= -m^2 \\ &= -E^2 + a^2\,(1+2\phi)\sum_i {P^i}^2 \end{aligned} \tag{6.26}$$

Writing $E^2 - m^2 = p^2$, where p is the magnitude of the momentum measured by the comoving observer, and noting that $\left(\sum_i {P^i}^2\right)^{1/2}$ is the magnitude of the 3-

momentum P^i, we get

$$\begin{aligned}\sum_i P^{i2} &= \frac{E^2 - m^2}{a^2(1+2\phi)}\\ \left(\sum_i P^{i2}\right)^{1/2} &= \frac{p(1-\phi)}{a}.\end{aligned} \tag{6.27}$$

We can write, as for any 3-vector, the 3-momentum as a product of its magnitude and direction,

$$P^i = \left(\sum_i P^{i2}\right)^{1/2} \hat{p}^i = \frac{p(1-\phi)}{a}\hat{p}^i, \tag{6.28}$$

where $\hat{p}^i$ is the unit vector in the direction of P^i, $\hat{p}^i\hat{p}^j\delta_{ij} = 1$.

To find $\mathrm{d}p/\mathrm{d}\lambda$ we must solve the geodesic equation.

Exercise 22

We can relate the derivatve w.r.t. p to the derivatve w.r.t. E using the relation $E^2 = p^2 + m^2$. The derivative w.r.t. E is obtained from the geodesic equation for P^0 and Eq. 6.25. We will need the Christoffel symbols $\Gamma^0_{\alpha\beta}$, defined in Eq. 1.53 for the perturbed metric, Eq. 6.15. Show that (with time derivative represented by an over-dot)

$$\begin{aligned}\Gamma^0_{00} &= \dot{\psi}\\ \Gamma^0_{0i} &= \psi_{,i}\\ \Gamma^0_{ij} &= a^2\delta_{ij}\left[H + 2H(\phi-\psi) + \dot{\phi}\right]\end{aligned} \tag{6.29}$$

Evaluate the geodesic equation for P^0 to show that

$$\frac{\mathrm{d}p}{\mathrm{d}\lambda} = \frac{-E^2}{a}\hat{p}^i\frac{\partial\psi}{\partial x^i} - Ep\left(H - H\psi + \dot{\phi}\right) \tag{6.30}$$

This equation is valid upto linear order and in particular includes the zeroth order terms. Upto first order, we therefore have,

$$\frac{1}{P^0}\frac{\mathrm{d}p}{\mathrm{d}\lambda} = -\frac{E}{a}\hat{p}^i\psi_{,i} - p\left(H + \dot{\phi}\right) \tag{6.33}$$

The last term in Eq. 6.23 is a second order term, since at zeroth order (an isotropic homogeneous Universe) the occupation number does not depend on $\hat{\mathbf{p}}$ and the direction of momentum of particles does not change along the geodesics. Therefore both $\partial f/\partial \hat{p}^i$ and $\partial \hat{p}^i/\partial \lambda$ are first order terms making the product a second order term.

Putting everything together we obtain for the left side of the Boltzmann equation 6.21 (see Eq. 6.23)

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}\lambda} &= P^0\frac{\mathrm{d}f}{\mathrm{d}t}\\ &= P^0\left[\frac{\partial f}{\partial t} + \frac{p}{E}\frac{\hat{p}^i\left(1-\phi+\psi\right)}{a}\frac{\partial f}{\partial x^i} + \frac{\partial f}{\partial p}\left(-\frac{E}{a}\hat{p}^i\psi_{,i} - p\dot{\phi} - pH\right)\right]\\ &= P^0\left[\frac{\partial f}{\partial t} + \frac{p}{E}\frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} - \frac{\partial f}{\partial p}\left(\frac{E}{a}\hat{p}^i\psi_{,i} + p\dot{\phi} + pH\right)\right] \end{aligned} \tag{6.34}$$

Note that the above expression is general and applies to all species.

6.2.2 Cold dark matter

In the hydrodynamic limit, we are interested in scales much larger compared to the mean free path of the particles. In this limit, we are not interested in the microscopic motions of individual particles. The hydrodynamic limit is therefore obtained by integrating the Boltzmann equation over momentum.

For cold dark matter there are no interactions, $C(f) = 0$, and it is cold implying that it has negligible kinetic energy and $E = m + p^2/(2m) \approx m$ The number density of dark matter, n_{cdm}, is given by integrating its occupation number over momentum space,

$$n_{\mathrm{cdm}}(\mathbf{x},t) = \int \frac{\mathrm{d}^3 p}{(2\pi)^3} f(\boldsymbol{p},\mathbf{x},t). \tag{6.35}$$

We note that the above integral is a local integral, to be evaluated at every spacetime point $(\mathbf{x}, t)$. In addition to microscopic thermal motions (due to negligible but non-zero temperature), dark matter may also have bulk flows or peculiar velocities, v^i_{cdm}. This is just the average bulk velocity of dark matter and is given by the first moment of the occupation number,

$$\begin{aligned} v^i_{\mathrm{cdm}} &= \frac{\int \frac{\mathrm{d}^3 p}{(2\pi)^3} f(\boldsymbol{p},\mathbf{x},t)\frac{p^i}{E}}{\int \frac{\mathrm{d}^3 p}{(2\pi)^3} f(\boldsymbol{p},\mathbf{x},t)}\\ &= \frac{1}{n_{\mathrm{cdm}}}\int \frac{\mathrm{d}^3 p}{(2\pi)^3} f(\boldsymbol{p},\mathbf{x},t)\frac{p^i}{E} \end{aligned} \tag{6.36}$$

In addition to integrals of occupation number over momentum and its first moment, looking at Eq. 6.34, we will also need to integrated the derivatives of the particle distributions at appropriate orders. At zeroth order the dark matter distribution is isotropic and therefore the integral over the momentum direction must vanish,

$$\int \frac{\partial f}{\partial p}\hat{p}^i \frac{\mathrm{d}^3 p}{(2\pi)^3} = 0, \tag{6.37}$$

since $f(p)$ and $\frac{\partial f}{\partial p}$ are independent of direction of momentum $\hat{p}^i$ and only depend on the magnitude p (e.g. the isotropic Bose-Einstein or Fermi-Dirac distribution). We also have, on integrating by parts,

$$\begin{aligned}\int \frac{\mathrm{d}^3 p}{(2\pi)^3} p \frac{\partial f}{\partial p} &= \int \frac{\mathrm{d}\Omega_p}{(2\pi)^3} \int p^3 \mathrm{d}p \frac{\partial f}{\partial p} \\ &= \int \frac{\mathrm{d}\Omega_p}{(2\pi)^3} \left(- \int 3p^2 f(p) \mathrm{d}p + \left[p^3 f\right]_0^\infty \right) \\ &= -3 \int \frac{\mathrm{d}^3 p}{(2\pi)^3} f + 0 \\ &= -3 n_{\mathrm{cdm}}, \end{aligned} \tag{6.38}$$

where we have assumed that the boundary term after integrating by parts in the second line, $p^3 f$, vanishes for physical distributions, since $f \to 0$ exponentially fast for $p \to 0$ and it rises slower than p^3 as $p \to 0$ (e.g. for Bose-Einstein distribution $f(p) \propto 1/p$ as $p \to 0$.

Integrating the Boltzmann equation for CDM over momenum and using the above results for integrals of different terms in Eq. 6.34 we get

$$\begin{aligned} 0 &= \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{\mathrm{d}f}{\mathrm{d}\lambda} \\ &= \frac{\partial n_{\mathrm{cdm}}}{\partial t} + \frac{n_{\mathrm{cdm}}}{a} \nabla \cdot \upsilon_{\mathrm{cdm}} + 3 n_{\mathrm{cdm}} \left(H + \dot{\phi}\right) \end{aligned} \tag{6.39}$$

up to first order. We have moved n_{cdm} out of the divergence in the second term in last line, since υ_{cdm} is already first order and we therefore need only the zeroth order part of n_{cdm} which has a vanishing gradient. The difference between $\nabla \cdot n_{\mathrm{cdm}} \upsilon_{\mathrm{cdm}}$ and $n_{\mathrm{cdm}} \nabla \cdot \upsilon_{\mathrm{cdm}}$ is a second order term. Note that the differentiation w.r.t $\mathbf{x}$ in the second term in Eq. 6.34 commutes with integration w.r.t $\boldsymbol{p}$, since they are independent variables and using Eq. 6.36 we get the second term in Eq. 6.39. Writing the dark matter number density as average or zeroth order number density plus a perturbation, $n_{\mathrm{cdm}} = \bar{n}_{\mathrm{cdm}} + \delta n_{\mathrm{cdm}}$, and substituting in Eq. 6.39, the zeroth order part gives us the usual zeroth order continuity equation implying $\bar{n}_{\mathrm{cdm}} \propto a^{-3}$

and for the first order continuity equation we get,

$$\frac{\partial \delta n_{\rm cdm}}{\partial t} + \frac{\bar{n}_{\rm cdm}}{a}\nabla \cdot \upsilon_{\rm cdm} + 3\delta n_{\rm cdm} H + 3\bar{n}_{\rm cdm}\dot{\phi} = 0. \tag{6.40}$$

Defining fraction perturbation, $\delta_{\rm cdm} = \delta n_{\rm cdm}/\bar{n}_{\rm cdm}$, we have

$$\frac{\partial \delta_{\rm cdm}}{\partial t} = \frac{1}{\bar{n}_{\rm cdm}}\frac{\partial \delta n_{\rm cdm}}{\partial t} + 3H\frac{\delta n_{\rm cdm}}{\bar{n}_{\rm cdm}}. \tag{6.41}$$

The continuity equation for the fractional dark matter pertubation therefore becomes

$$\frac{\partial \delta_{\rm cdm}}{\partial t} + \frac{1}{a}\nabla \cdot \upsilon_{\rm cdm} + 3\dot{\phi} = 0. \tag{6.42}$$

Compared to the Newtonian continuity equation (Eq. 5.34), we have an extra term ($3\dot{\phi}$) in relativistic hydrodynamics. There is an obvious reason for the presence of this term. The potential ϕ is the perturbation to the spatial part of the metric which tells us how to measure distances and volumes. This extra term in the continuity equation tells us that in addition to matter flowing in or out of a volume, the density can also change because the metric perturbations or gravity changes changing the measure of volume. This is the pertubed version of change in density due to expansion of Universe at zeroth order. At zeroth order, the zeroth order part of the metric $\propto a$ evolves and changes the volume and hence density. At first order, the peturbation to the volume is represented by the corresponding metric perturbation ϕ. The $\dot{\phi}$ term here is therefore analogous to the $\dot{a}$ or H term in zeroth order continuity equation, Eq. 5.24.

To get the relativistic momentum conservation or Euler equation in curved space-time, we take the first moment of the Boltzmann equation to get,

$$\frac{1}{\bar{n}_{\rm cdm}}\int \frac{{\rm d}^3 p}{(2\pi)^3}\frac{p\hat{p}^i}{E}\frac{{\rm d}f}{{\rm d}\lambda} = 0 \tag{6.43}$$

For the Euler equation, we need the integral in the following exercise.

Exercise 23

Show that

$$\int {\rm d}\Omega_p \hat{p}^i \hat{p}^j = \frac{4\pi}{3}\delta^{ij}, \tag{6.44}$$

where ${\rm d}\Omega_p$ is the angular part of ${\rm d}^3 p$. Hint: Choose a coordiante system.

The second term in the Euler equation is of order $p^2/E^2 \sim v^2$ multiplied with $\partial f/\partial x^i$ and can be ignored. On integration, it will give terms of order $v_{\rm cdm}^2$, i.e. second order. We also need to evaluate

$$\begin{aligned}
\int \frac{{\rm d}^3 p}{(2\pi)^3}\frac{p^2}{E}\frac{\partial f}{\partial p}\hat{p}^j &= \int \frac{{\rm d}\Omega_p}{(2\pi)^3}\hat{p}^j \int {\rm d}p \frac{p^4}{E}\frac{\partial f}{\partial p} \\
&= -\int {\rm d}p f \int \frac{{\rm d}\Omega_p}{(2\pi)^3}\hat{p}^j\left[\frac{4p^3}{E} - \frac{p^4}{E^2}\frac{{\rm d}E}{{\rm d}p}\right] \\
&= -4\int \frac{{\rm d}^3 p}{(2\pi)^3}\frac{p^j}{E} f = -4 n_{\rm cdm} v_{\rm cdm}^j, \qquad (6.45)
\end{aligned}$$

where we have neglected the $p^4/E^2({\rm d}E/{\rm d}p) = p^5/E^3 \sim v^3$ term.

Now we have all integrals involved in evaluation of the first moment of Boltzmann equation giving

$$\begin{aligned}
0 &= \frac{1}{n_{\rm cdm}}\frac{\partial\left(v_{\rm cdm}^j n_{\rm cdm}\right)}{\partial t} + 4v_{\rm cdm}^j\left(H + \dot{\phi}\right) + \frac{3}{a}\frac{1}{3}\psi^{,j} \\
&= \frac{\partial v_{\rm cdm}^j}{\partial t} + v_{\rm cdm}^j H + \frac{1}{a}\psi^{,j}. \qquad (6.46)
\end{aligned}$$

This is identical to the Newtonian version , Eq. 5.31, with the pressure term neglected.

The only difference for dark matter between Newtonian and relativistic treatment is the extra $\dot{\phi}$ term in the continuity equation. It turns out that this term is important only during transition from an expansion dominated by one type of equation of state to another, for example during the transition from radiation domination to matter domination or from matter domination to dark energy domination. In particular this term can be neglected when background expansion is dominated by one type of fluid. Therefore the solutions for growth of dark matter perturbations that we derived for radiation dominated, matter dominated and dark energy dominated eras hold in the relativistic cosmology also.

We will be using conformal time η as our time variable from now on and denote a derivative w.r.t. η by a prime ($'$). Also, we will work in Fourier space, with $\nabla \rightarrow i\mathbf{k}$, where $\mathbf{k}$ is the comoving wavenumber. Also we are interested in scalar modes, for which $v_{\rm cdm}^j = v_{\rm cdm}\hat{k}^j$. The dark matter continuity and Euler equations, Eqs. 6.42 and 6.46. in Fourier space are

$$\begin{aligned}
\delta_{\rm cdm}' + ikv_{\rm cdm} + 3\phi' &= 0 \\
v_{\rm cdm}' + v_{\rm cdm}\frac{a'}{a} + ik\psi &= 0 \qquad (6.47)
\end{aligned}$$

Chapter 7

Cosmic microwave background: perturbations

We get an equation of continuity for the CMB energy density perturbations, $\delta_\gamma = (\rho_\gamma - \bar{\rho}_\gamma)/\bar{\rho}_\gamma$, that is similar to that for dark matter,

$$\frac{\partial \delta_\gamma}{\partial t} + \frac{1}{a}\nabla \cdot \upsilon_\gamma + 4\dot{\phi} = 0, \tag{7.1}$$

with the factor of 4 in the last term taking into account that the metric perturbation in addition to changing the volume also affect the wavelength of photons. δ_γ included perturbation to the number density of photons due to change in volume as well photon energy. The photon bulk *velocity* υ_γ is defined in a way similar to the dark matter as first moment of the photon distribution function,

$$\upsilon_\gamma^j = \frac{1}{\bar{n}_\gamma}\int \frac{\mathrm{d}^3 p}{(2\pi)^3}\hat{p}^j f(\boldsymbol{p}, \mathbf{x}, t), \tag{7.2}$$

where $\bar{n}_\gamma$ is the avearage number density of CMB photons and we have used $\boldsymbol{p}/E = \hat{\mathbf{p}}$ for photons.

It is more convenient, for photons, to work with temperature perturbations instead of energy density perturbations. As long as the CMB has a blackbody spectrum locally, it can be fully described by its temperature, T. We therefore define

$$\Theta(\hat{\mathbf{p}}, \mathbf{x}, t) \equiv \frac{\Delta T}{T}(\hat{\mathbf{p}}, \mathbf{x}, t). \tag{7.3}$$

Note that the temperature pertubations still depend on the line of sight direction $\hat{\mathbf{p}}$ from which the photons are coming, since we can have different temperature for the photons coming from different directions. The dependence on the photon energy,

p drops out. Thus, we have a Planck spectrum for photons everywhere, since the early Universe was in *local thermal equilbrium* everywhere. However, since there were energy density fluctuations, the temperature of photons is perturbed and varies from point to point. We can see the relation between the perturbations of the occupation number and temperature more rigorously as follows. We are interested in linear perturbations, therefore we can write the occupation number of any species $f(p, \hat{\mathbf{p}}, bx, t)$ as a zeroth order or background occupation number, $f_0(p, t)$, and a perturbation, $f_1(p, \hat{\mathbf{p}}, \mathbf{x}, t)$. For photons, the spectrum is given by the Planck spectrum and therefore,

$$f(p, \hat{\mathbf{p}}, \mathbf{x}, t) = \frac{1}{e^{\frac{hp}{k_B T(\hat{\mathbf{p}}, \mathbf{x}, t)}} - 1}. \tag{7.4}$$

We can write the temperature T as a zeroth order average temperature, $\bar{T}$ plus a perturbation,

$$T = \bar{T} + \Delta T(\hat{\mathbf{p}}, \mathbf{x}, t) = \bar{T}\left[1 + \Theta(\hat{\mathbf{p}}, \mathbf{x}, t)\right] \tag{7.5}$$

The perturbation in the occupation number, f_1 is related to the perturbation in temperature by Taylor expansion of the occupation number f in T around average temperature $\bar{T}$. The first order term in the Taylor seried expansion is

$$f_1(p, \hat{\mathbf{p}}, \mathbf{x}, t) = \frac{\partial f}{\partial T}\Delta T = \frac{-p}{T}\frac{\partial f}{\partial p}\Delta T = -p\Theta\frac{\partial f}{\partial p} \tag{7.6}$$

We can expand Θ in spherical harmonics, resloving the dependence on angular direction $\hat{\mathbf{p}}$ of photons into dependence on angular quantum numbers ℓ, m, with the transformations from one space to the other given by

$$\begin{aligned} \Theta(\mathbf{x}, \hat{\mathbf{p}}, t) &= \sum_{\ell m} a_{\ell m}(\mathbf{x}, t) Y_{\ell m}(\hat{\mathbf{p}}) \\ a_{\ell m} &= \int \mathrm{d}\Omega \Theta(\hat{\mathbf{p}}, \mathbf{x}, t) Y^*_{\ell m}(\hat{\mathbf{p}}) \end{aligned} \tag{7.7}$$

where Ω is the solid angle and

$$\int \mathrm{d}\Omega \cdots \equiv \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \sin\theta \mathrm{d}\theta \ldots, \tag{7.8}$$

an * represents complex conjugation and the spherical harmonics satisfy the orthogonality and completeness relations,

$$\int \mathrm{d}\Omega Y_{\ell_1 m_1}(\theta, \phi) Y^*_{\ell_2 m_2}(\theta, \phi) = \delta_{\ell_1 \ell_2}\delta_{m_1 m_2}. \tag{7.9}$$

A good reference for spherical harmonics, their properties and useful relations is [49].

The spherical harmonics consitute a complete set of orthogonal functions on a sphere, and are analogous to Fourier transform on Euclidean space. The orbital quantum number ℓ is analogous to the magnitude of the Fourier wave number $k = |\mathbf{k}|$, and represents the scale of the perturbation. The other quantum number m is analogous to the direction of the Fourier mode $\hat{\mathbf{k}}$ and has information about the directional dependence of the perturbation.

We can however make further simplifications. We can take the usual Fourier transform w.r.t the spatial dependence $\mathbf{x}$ to get the Fourier space temperature perturbation $\Theta(\mathbf{k}, \hat{\mathbf{p}}, t)$.

For a *statistically* isotropic density and temperature field and isotropic physics, i.e. there is nothing to single out a particle direction in the Universe and any directional dependence or anisotropy is random, the evolution equations or equations of motion should not depend on m or $\hat{\mathbf{p}}$ and perturbation mode direction $\hat{\mathbf{k}}$ but only the scale $k \equiv |\mathbf{k}|$ or ℓ. This implies that the directions $\hat{\mathbf{k}}$ and $\hat{\mathbf{p}}$ cannot appear alone in the evolution equations but only as a scalar combination $\mu \equiv \hat{\mathbf{k}} \cdot \hat{\mathbf{p}}$, since evolution can and does depend on the relative direction between where the photons are going and the direction of Fourier mode. All of these properties emerge automatically when we derive the Botzmann equation for photons rigorously, e.g. see [47]. Thus we have,

$$\Theta(\mathbf{k}, \hat{\mathbf{p}}, t) \equiv \Theta(k, \hat{\mathbf{k}} \cdot \hat{\mathbf{p}}, t) \equiv \Theta(k, \mu, t), \tag{7.10}$$

where μ is the cosine of the angle between $\mathbf{k}$ and $\boldsymbol{p}$. Thus the evolution of cosmological perturbations depends only on the relative direction between $\hat{\mathbf{p}}$ and $\hat{\mathbf{k}}$ but not on absolute directions. Another way to understand this is that the only reference direction w.r.t which to measure the direction in which the photons are travelling is provided by the only other vector present in the homogeneous and isotropic Universe and that is $\hat{\mathbf{k}}$. Since all our evolution equations are linear in perturbation variables, different $\mathbf{k}$ modes do not couple to each other and evolve independently. Thus for each $\mathbf{k}$ mode we are free to choose a coordiante system, and we can in particular choose the $\hat{\mathbf{z}}$ axes along the $\mathbf{k}$ and thus $\hat{\mathbf{p}}$ dependence just translates into the dependence on the cosine of the polar angle $\mu = \cos\theta$.

Since $\Theta(k, \mu, t)$ is now only a function of μ, it can be expanded in Legendre polynomials (P_ℓ), which are just the spherical harmonics with their directional dependence averaged out,

$$P_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{p}}) = \frac{4\pi}{2\ell + 1} \sum_m Y_{\ell m}(\hat{\mathbf{k}}) Y^*_{\ell m}(\hat{\mathbf{p}}) \tag{7.11}$$

The decomposition of temperature anisotropies into multipoles Θ_ℓ is given by the relations

$$\Theta(k,\mu,t) = \sum_\ell (-i)^\ell \,(2\ell+1)\, P_\ell(\mu)\Theta_\ell(k,t)$$
$$\Theta_\ell(k,t) = \frac{1}{(-i)^\ell}\int_{-1}^{1}\Theta(k,\mu,t)P_\ell(\mu)\frac{\mathrm{d}\mu}{2} \tag{7.12}$$

From above definitions, we see that the monopole Θ_0 does not involve any factor of $\hat{\mathbf{p}}$ and thus corresponds to the energy density perturbation, δ_γ, with the relation

$$\delta_\gamma = 4\Theta_0 \tag{7.13}$$

following from $\rho_\gamma \propto T^4$. Also, Θ_1 involves one factor of $\hat{\mathbf{p}}$ and thus corresponds to the velocity perturbation, υ_γ.

Exercise 24

Find relation between υ_γ and Θ_1. Hint: Take dot product of Eq. 7.2 with $\hat{\mathbf{k}}$ and use definition of Θ in Eq. 7.6 and the fact that for scalar modes $\upsilon_\gamma^j = \upsilon_\gamma \hat{k}^j$.

The quadrupole, Θ_2 involves second moment of momentum and thus corresponds to the anisotropic off-diagonal space-space components of the stress-energy tensor. For an ideal fluid the quadrupole and higher multipole, $\ell \geq 2$, components vanish. However the photon-baryon plasma is not an ideal fluid, and when we take into account finite mean free path of photons in the plasma, we have non-negligible shear viscosity and therefore non-zero quadrupole. After recombination photons free stream resulting in complete breakdown of the fluid approximation and we must evolve the infinite hierarchy of multipoles. Today, there is non-negligible CMB multipoles upto $\ell \approx 2500$ which have been measured by space experiment Planck [50] as well as ground based experiments on the south pole [51] and Atacama desert in Chile [52].

Doing a calculation similar to that for dark matter, we get the evolution equation for baryons (see [47] for derivation of collision term as well as evolution equations of photons). The equations for the two non-relativistic species as almost identical, the only difference from the CDM equations, Eq. 6.47, being an extra collision term for taking into account the Thomson scattering between photons and electrons,

$$\delta_{\mathrm{b}}' + ik\upsilon_{\mathrm{b}} + 3\phi' = 0$$
$$\upsilon_{\mathrm{b}}' + \upsilon_{\mathrm{b}}\frac{a'}{a} + ik\psi = \frac{\tau_{\mathrm{T}}'}{R}\left(\upsilon_{\mathrm{b}} + 3i\Theta_1\right), \tag{7.14}$$

where we differential Thomson optical depth is defined as (Eq. 4.21)

$$\frac{d\tau_T}{d\eta} \equiv \tau'_T = -n_e\sigma_T a \tag{7.15}$$

$$\tau_T(\eta) = \int_{\eta}^{\eta_0} n_e\sigma_T a d\eta, \tag{7.16}$$

where η_0 is the conformal time today and

$$R = \frac{3}{4}\frac{\bar{\rho}_b}{\bar{\rho}_\gamma} \tag{7.17}$$

The evolution equations for photons, in terms of temperature perturbations Θ, are

$$\Theta'_0 + k\Theta_1 = -\phi' \tag{7.18}$$

$$\Theta'_1 - \frac{k\Theta_0}{3} = \frac{k\psi}{3} + \tau'_T\left(\Theta_1 - \frac{iv_b}{3}\right) \tag{7.19}$$

Note that the collision terms in the baryon and photon equations are identical as required by conservation of momentum. The collision term transfers momentum between the photons and baryons but the momentum gained by the photons (baryons) must be equal to the momentum lost by the baryons (photons) remains conserved. The factor of R takes into account the difference in energy densities of baryons and photons and ensures momentum conservation. At high redshifts, during radiation domination, $R \propto 1/(1+z)$, is very small and $\tau'_T \propto n_e$ is very large. This suggests a perturbative solution in the small quantity R/τ'_T that we will derive in the next section. Note that we are still working in the Thomson limit and ignoring the change in the *magnitude* of momentum of photon in individual Thomson scatterings in the electron rest frame. However, the change in the direction of photons in the Thomson scattering is enough to provide a radiation drag force and couple baryons and photons. At first order, the change in energy of photons due to Doppler shift from moving electrons is included at first order in velocity of electrons. The $k\Theta_0/3$ term is just the gradient of radiation pressure with the sound speed for radiation ${c_s}^2 = 1/3$. It can be compared with similar pressure terms in Euler equation for non-relativistic fluids.

7.1 Tight coupling solutions: acoustic oscillations

We will try to get some analytic insights into the phyiscs of CMB ansiotropies. The analytic approach we follow here is partly based on [53, 54] We can rewrite the baryon velocity equation, Eq. 7.14, as

$$v_b = -3i\Theta_1 + \frac{R}{\tau'_T}\left(v_b{}' + v_b\frac{a'}{a} + ik\psi\right). \tag{7.20}$$

When the rate of Thomson scattering of photons with electrons, τ'_T is much larger compared to the Hubble expansion rate, the photons and baryons are tightly coupled together. In particular, they move together, since photons are trapped in the plasma with very small mean free path. The baryons and photons acts as a single fluid in the tight coupling limit. We see from Eq. 7.20 that in the tight coupling approximation, $\tau'_T \to \infty$, terms in the brackets can be neglected and the baryon and photon velocities coincide,

$$v_b = -3i\Theta_1 \tag{7.21}$$

In reality, τ'_T is large but finite. We can take into account the difference between the baryon and photon velocties by solving Eq. 7.20 iteratively. At zeroth order in R/τ'_T, we have the solution Eq. 7.21. We can find the next order term in R/τ'_T by substituting the zeroth order solution in the right hand side of Eq. 7.20,

$$v_b = -3i\Theta_1 + \frac{R}{\tau'_T}\left(-3i\Theta'_1 - 3i\Theta_1\frac{a'}{a} + ik\psi\right). \tag{7.22}$$

We can use this solution to eliminate v_b from the Eq. 7.19 for Θ_1 and then combine Eqs. 7.18 and 7.19 to eleminate Θ_1 also and get a single second order equation Θ_0,

$$\Theta''_0 + \frac{a'}{a}\frac{R}{1+R}\Theta'_0 + k^2{c_s}^2\Theta_0 = F(\phi, \psi, R), \tag{7.23}$$

where

$$c_s = \sqrt{\frac{1}{3(1+R)}} \tag{7.24}$$

is the sound speed of the coupled baryon-photon fluid. At high redshifts, during radiation domination, $R \approx 0$ and $c_s \approx 1/\sqrt{3}$.

Exercise 25

Derive Eq. 7.23 and show that

$$F(\phi, \psi, R) = \frac{-k^2\psi}{3} - \frac{a'}{a}\frac{R}{1+R}\phi' - \phi'' \tag{7.25}$$

Eq. 7.23 is an equation for a *damped forced* oscillator. The damping is provided by the Hubble term ($\propto \Theta_0'$) and $F(\phi, \psi, R)$ is the forcing term and is sourced by the gravity of all fluid components, including baryons, photons, neutrinos and dark matter.

7.1.1 Case I: No gravity

Lets try to understand this equation, which describes the dynamics of acoustic oscillations in the primordial plasma, by analyzing the effect of different terms. First we will ignore the damping and forcing terms to get the simplest limit of Eq. 7.23

$$\Theta_0'' + k^2 c_s(\eta)^2 \Theta_0 = 0, \tag{7.26}$$

where we have made explicit the dependence on time of the sound speed c_s which decreases from the ultrarelativistic limit of $1/\sqrt{3}$ as we go from radiation domination towards matter domination. The solution is given by

$$\Theta_0 = A\cos(kr_s) + B\sin(kr_s), \tag{7.27}$$

where

$$r_s = \int_0^\eta c_s \mathrm{d}\eta \tag{7.28}$$

is the sound horizon. At high redshifts or early times, $c_s \approx 1/\sqrt{3}$. For the adiabatic initial conditions, the energy denstiy and curvature perturbations are frozen in and constant outside the horizon, as $kr_s \approx k\eta/\sqrt{3} \to 0$. This implies that the second term $\propto k\eta$ as $k\eta \to 0$ must vanish, and $B = 0$. Thus adiabatic perturbations only excite the cosine modes of acoustic oscillations. The adiabatic initital or boundary conditions also imply, by vanishing of the sine mode, that the sound waves which are excited are standing sound waves. The full real space solution for a single mode is $\propto \cos(k\eta/\sqrt{3})e^{i\mathbf{k}.\mathbf{x}} = \cos(\omega\eta)e^{i\mathbf{k}.\mathbf{x}}$ as opposed to a travelling wave solution which should be $e^{i\mathbf{k}.\mathbf{x}\pm\eta\omega}$, where $\omega = k/\sqrt{3}$ is the angular frequency of the wave. Thus in order to have a travelling wave solution, we would have needed $B = \pm iA$. Remember that Θ_0 is the amplitude of the Fourier mode with wavenumber k, since after substituting the spatial plane wave solutions we are solving for the amplitude of the spatial Fourier modes. This amplitude of the standing wave of fixed wavenumber oscillates with frequency $\omega =\approx k/\sqrt{3}$, as expected for a standing wave, similar to the oscillations of a string fixed at both ends. From Eq. 7.18, we get the solution for Θ_1 (ignoring gravity)

$$\Theta_1 = -\frac{\Theta_0'}{k} = A\sin(kr_s){r_s}'$$
$$= Ac_s\sin(kr_s). \tag{7.29}$$

We therefore have standing sound waves with the internal energy density of the sound waves $\propto \Theta_0^2$ and the kinetic energy $\propto \Theta_1^2$ or the square of the fluid velocity. Note that the fluid velocities, in particular the velocity of the photon fluid, $v_\gamma \propto \Theta_1$, is still non-relativistic even though the fluid is made up of the massless photons bouncing around in the plasma. As in soundwaves in usual non-relativistic fluids, the kinteic energy and internal energy are out of phase and oscillate into each other,.The total energy in the sound wave in a non-relatvistic fluid [55] with average mass density ρ, mass density perturbation $\delta\rho$ and veloicty perturbation v is $\frac{1}{2}\delta\rho {c_s}^2 + \frac{1}{2}\rho v^2$, where first term represents the thermal or internal energy and the second term the kinetic energy of the wave. For our ultrarelativistic fluid, during radiation domination most of the energy is in the photonsand the internal energy $\propto {c_s}^2\Theta_0^2$ while the kinetic energy is proportional to Θ_1^2 [56] giving the total energy $\propto {c_s}^2\Theta_0^2 + \Theta_1^2 = A^2{c_s}^2\left(\cos^2 kr_s + \sin^2 kr_s\right) = A^2{c_s}^2$. We should keep in mind that energy density is not a frame independent quantity and it will be different in a different gauge.

In fact, the temperature of the CMB and its moments Θ_0 and Θ_1 are also gauge dependent. We are interested what we will observe which should not depend on the gauge we choose for calculations. In particular, the $\ell \geq 2$ multipoles are gauge invariant (similarly to the tensor modes) and these are the multipoles that we observe today and which are the foundation for the standard cosmological model. We will see later that $\ell \geq 2$ modes that we observe today are sourced by gauge invariant combinations of the perturbations in fluid and metric at the time of recombination. The $\ell \geq 2$ modes are gauge invariant at first order in perturbations since changes in local frame, such as gravitational potentials or velocity boosts, are $\ell = 0, 1$ terms and cannot affect $\ell \geq 2$ modes by the orthogonality of Legendre polynomials.

For now, lets us assume that the recombination happens instantly at time η_*. At $\eta < \eta_*$ we have standing acoustic oscillations in the baryon-photon plasma excited by initial scalar perturbations and the amplitudes of different k modes will be in different phases of oscillation at $\eta = \eta_*$. In particular the modes for which $kr_s(\eta_*) \approx k\eta_*/\sqrt{3} = n\pi$, where $n \geq 1$ is an integer, would have the maximum amplitude (for Θ_0) at the time of recombination while modes with $kr_s(\eta_*) = (2n-1)\pi/2$ modes would be have the minimum amplitude. At $\eta = \eta_*$ the Universe recombines and (almost) all electrons get bound in neutral atoms. The Universe thus goes from opaque to transparent for the CMB photons. The acoustic oscillations stop and the CMB photons carry the image of the Universe, in particular the state of the plasma and the phases of oscillations of different k modes at the last scattering surface $\eta = \eta_*$, along the null geodesics from η_* to today $\eta = \eta_0$. What we see today is a spherical projection or cut of 3-D acoustic oscillations at η_*. The variations in CMB temperature, $\Theta_0 \propto \cos(kr_s(\eta_*))$, on the last scattering surface appear to us variations in temperature of CMB along different directions. The $\ell = 0$ mode of the CMB anisotropies that we observe has contribution from graviational potentials at our position while $\ell = 1$ has contribution from our motion w.r.t CMB. The $\ell \geq 2$ modes are not affected by the metric perturbations or our frame of reference. The temperature anisotropies we observe in $\ell \geq 2$ modes today is not just Θ_0 that we have calculated in conformal Newtonian gauge and which will be different if we chose a different gauge. What we observe is the gauge invariant combination $\Theta_0 + \psi$. This can be thought of as the temperature perturbation in the conformal Newtonian gauge modifed by the gravitational redshift that the CMB photons experience as they free stream to us from regions on the last scattering surface with different gravitational potentials. We will see that there are other important terms also which contribute and modify the CMB ansiotropies that we observe but $\Theta_0 + \psi$ term is the dominant contribution and we will focus on it for now.

7.1.2 Case II: gravity dominated by dark matter

Note that gravity has so far played no role in the solutions 7.27 and7.29. These are pure sound waves driven by radiation pressure. Lets now introduce gravity in simplest way possible. Let us assume that the mertic perturbations, ϕ, ψ are constant (e.g. in matter dominated regime they are sourced dominantly by the dark matter). We can thus ignore the ϕ' and ϕ'' terms in Eq. 7.23 and we are left with a forcing term that is constant,

$$\Theta_0'' + k^2 {c_s}^2 \Theta_0 = \frac{-k^2\psi}{3}, \tag{7.30}$$

where we are still ignoring the damping term i.e. negligible baryon density $R \to 0$ so that dark matter gives the dominant contribution to ψ. A constant forcing term just shifts the zero point of the harmonic oscillator and the solution to Eq. 7.30 is given by

$$\Theta_0 = (\Theta_0(0) + \psi(0)) \cos\left(\frac{k\eta}{\sqrt{3}}\right) - \psi, \tag{7.31}$$

where $\Theta_0(0)$ and $\psi(0)$ are the initial conditions at $\eta = 0$. For the adiabatic initial conditions, we have the following relation between different initial perturbative variables in the radiation dominated era:

$$\Theta_0(0) = \frac{-\psi(0)}{2} = \frac{\phi(0)}{2} = \frac{\delta_{\rm cdm}(0)}{3} = \frac{\delta_{\rm b}(0)}{3} = \frac{\delta_\gamma(0)}{4} \tag{7.32}$$

Therefore, for $\Theta_0(0) = A$, we have $\psi(0) = -2A$ and $\delta_{\rm cdm}(0) = 3A$. The solution therefore becomes,

$$\Theta_0 = -A\cos\left(\frac{k\eta}{\sqrt{3}}\right) + 2A. \tag{7.33}$$

Thus the effect of the gravity of dark matter is to give a 180^o phase shift to the acoustic oscillations in addition to a constant shift [1]. As discussed above, taking into account the gravitational redshift as the CMB photons escape the gravitational potentials at the last scattering surface, the temperature fluctuations observed by us will be given by

$$\Theta_0(\eta_*) + \psi(\eta_*) = -A\cos\left(\frac{k\eta}{\sqrt{3}}\right), \tag{7.34}$$

since $\psi(\eta_*) = \psi(0) = -2A$. Thus the gravitational redshift from ψ exactly cancels the shift in the zero point of the oscillations. The 180^o phase shift however remains compared to the solution in the absence of any gravity.

We can understand the phase-shift as follows. Suppose we look at a region which had an initial overdensity, i.e. for A positive, we look at the region which corresponds to the peak of the initial perturbation in a particular mode k. In the absence of gravity, this region will also correspond to a region of peak pressure and the pressure in this region would have driven the density this region to decrease. In the presence of gravity however, it turns out that the gravity of dark matter is more than enough to overcome this pressure and instead of decompressing, the fluid in the peak region is compressed even more under the action of gravity until the raditaion pressure becomes strong enough to prevent further compression. This point now becomes the new maxima of the oscillation and happens at $k\eta/\sqrt{3} = \pi$, at a point where we would have had a minima in the absence of gravity. Thus we have a shift of size A in the maxima and the zero point. Once we take into account the redshift of size A, the zero point of the observed CMB temperature comes back to the original value, but phase-shift due to the gravity of dark matter remains. That the gravity of dark matter can overcome the radiation pressure is ensured by the adiabatic initial conditions.

[1] There is a small difference from [54] as they use the superhorizon solution during matter dominated era as initial condition while we use the superhorizon solution during the radiation dominated era

7.1.3 Case III: gravity dominated by baryons

Lets consider the other extreme case when there is negligible dark matter and gravity is dominated by baryons. In this case the driving force is also oscillating in phase with the acoustic oscillations. In the regions which are undergoing extreme compression, gravity is also enhanced and drives the compression even further and vice versa for underdensities. The synchronous driving force therefore amplifies the acoustic oscillations. Introducing baryons also decreases the sound speed, ${c_s}^2 = \frac{1}{3(1+R)}$, and hence makes the pressure term even weaker compared to gravity, increasing the amplitude of oscillations even further. Such extreme amplitudes of acoustic oscillations are ruled out by observations and are one of the strongest arguments against modification of newtonian dynamics (MOND) as an alternative to dark matter [57].

We can see the amplification using Eq. 7.30 but with ψ now not a constant but sourced by baryons and radiation. Approximating gravity by Poisson equation, with gravity dominated by radiation and baryons,

$$-k^2\psi = k^2\phi = 4\pi Ga^2\delta\rho = 4\pi Ga^2\left(\rho_b\delta_b + \rho_\gamma\delta_\gamma\right). \tag{7.35}$$

For tightly coupled baryons and photons, the perturbation in baryon density, δ_b is related to the photon energy density by adiabatic relation, i.e. fractional perturbations in number density of baryons is equal to the fractional perturbation in number density of photons which is three-fourths of the fractional perturbation in the energy density of photons,

$$\delta_b = \frac{3}{4}\delta_\gamma = 3\Theta_0. \tag{7.36}$$

The Poisson equation therefore becomes,

$$\begin{aligned}-k^2\psi &= 4\pi Ga^2\left(3\Theta_0\rho_b + 4\Theta_0\rho_\gamma\right)\\ &= 4\pi Ga^2\left(\rho_\gamma+\rho_b\right)\Theta_0\left(\frac{3\rho_b+4\rho_\gamma}{\rho_b+\rho_\gamma}\right)\\ &= \mathcal{H}^2 6\Theta_0\left(\frac{1+R}{1+4R/3}\right)\\ &\equiv 3\alpha^2\mathcal{H}^2\Theta_0,\end{aligned} \tag{7.37}$$

where we have defined the conformal Hubble parameter, $\mathcal{H}$,

$$\mathcal{H}^2 = \left(\frac{a'}{a}\right)^2 = \frac{8\pi Ga^2\rho}{3} \tag{7.38}$$

and dimensionless parameter

$$\alpha^2 = 2\left(\frac{1+R}{1+4R/3}\right). \tag{7.39}$$

In the limit $R \to 0$, $\alpha^2 \to 2$ and in the limit $R \gg 1$, $\alpha^2 \to 1.5$ giving $1.5 \leq \alpha^2 \leq 2$. The forced oscillator equation is therefore given by

$$\Theta_0'' + k^2 {c_s}^2 \Theta_0 = \alpha^2 \mathcal{H}^2 \Theta_0, \tag{7.40}$$

with solution

$$\Theta_0 \propto e^{\pm i\sqrt{k^2{c_s}^2 - \alpha^2\mathcal{H}^2}\eta}. \tag{7.41}$$

The modes which are of the size of horizon, $k\eta \sim \mathcal{H}$, we have $\alpha\mathcal{H} > kc_s$, since $\alpha^2/{c_s}^2 \gtrsim 3$ and the frequency is imaginary and the modes get enhanced as they enter the horizon until frequency becomes real again and acoustic oscillations start. Note that the initial conditions with initial velocity ~ 0 implies that the growing mode would be the solution that is picked up as matter falls into the gravitational wells. In other words, initially the Jeans scale is smaller than the mode size and the density perturbations are enhanced until the Jeans scale is larger than the mode size and we have acoustic oscillations. The frequency of oscillations is also decreased which results in shift in the peak positions on horizon entry. On timescale of order Hubble time, $\mathcal{H}\eta \sim 1$, we expect amplifications of order $e \sim 3$. Indeed, numerical calculations show that we can have quite large amplification of acoustic oscillations. In addition, in a no-dark matter Universe we would expect large acoustic oscillations in the matter power spectrum also, see [57] for an interesting discussion. Note that, we have so far ignored the friction term, which would damp the oscillations, though not exponentially. We therefore need a cold dark matter component to dominate the gravity so as to keep the oscillation amplitude at a level consistent with the observations.

7.1.4 Case IV: solution in presence of both baryons and dark matter

We can do a little better and take into account the presence of both baryons and cold dark matter. Notice the Θ_0 and ϕ terms appear in a very similar way in Eq. 7.23. We can therefore rewrite the equation by rearranging the terms as

$$\left(\frac{d^2}{d\eta^2} + \frac{R'}{1+R}\frac{d}{d\eta} + k^2{c_s}^2\right)(\Theta_0 + \phi) = \frac{k^2}{3}\left[\frac{1}{1+R}\phi - \psi\right], \tag{7.42}$$

where we have used

$$\frac{a'}{a}\frac{R}{1+R} = \frac{R'}{1+R}. \tag{7.43}$$

The equation simplifies with the approximation $R, \phi, \psi \approx$ *constant*, and $\phi = -\psi$,

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}\eta^2} + k^2{c_\mathrm{s}}^2\right)(\Theta_0 + \phi) = k^2{c_\mathrm{s}}^2\left[\phi - (1+R)\psi\right] = k^2{c_\mathrm{s}}^2\phi(2+R) \tag{7.44}$$

and the solution is

$$\Theta_0 + \phi = C\cos(kr_\mathrm{s}) + \phi(2+R). \tag{7.45}$$

we can find the constant C in terms of the initial conditions, $\Theta_0(0)$ and $\psi(0)$ by subsitituing $\eta = 0$ in the solution. We are interested in the observed temperature perturbation $\Theta_0 + \psi = \Theta_0 - \phi = \Theta_0 + \phi - 2\phi$, given by

$$\Theta_0 + \psi = \left[\Theta_0(0) + \psi(0)(1+R)\right]\cos(kr_\mathrm{s}) - \psi R, \tag{7.46}$$

where ψ has dominant contribution from cold dark matter in the matter dominated era (which is why ψ constant is a good approximation). Thus we see that the presence of both the baryons and the dark matter results in an offset ψR which is not completely cancelled. This offsets is in the same direction as the first peak and adds to it, with the amplitude of the first peak (and all odd peaks) given by (at $kr_\mathrm{s} = \pi$) $-\Theta_0(0) - \psi(0)(1+R) - \psi R = -A + 2A(1+R) + 2AR = A(1+4R)$ while the second extremum (and all even extrema) at $kr_\mathrm{s} = 2\pi$ gets reduced with amplitude given by $\Theta_0(0) + \psi(0)(1+R) - \psi R = \Theta_0(0) + \psi(0) = A$. The difference in the peaks is therefore proportional to the fraction of baryons in the Universe. It is important to note that in the absence of dark matter, the gravitational potentials would be provided by baryons and photons and the forcing term would oscillate in sync with acoustic waves making the current solution which assumes constant forcing term invalid. It is the presence of both the baryons and the dark matter which results in asymmetry in the odd and even extrema. The asymmetry depends on fraction of baryons, through R, as well as fraction of dark matter through ψ. We usually are interested in the power spectrum, which is the correct statistical quantity that can be compared between the theory and observations. The power spectrum is $\propto (\Theta_0 + \psi)^2$ and both the maxima and minima appear as peaks in the power spectrum while the zero crossings of the *transfer function* $\Theta_0 + \psi$ correspond to the minima of the power spectrum. Thus we have an odd-even asymmetry in the peaks of the CMB power spectrum as well as the baryon or matter power spectrum. The imprint of the acoustic oscillations in the matter power spectrum is called *baryon acoustic oscillations* or BAO.

7.1.5 Varying gravitational potentials - effect of free streaming neutrinos

We can also write a formal solution to Eq. 7.42,

$$\left(\frac{d^2}{d\eta^2}+\frac{R'}{1+R}\frac{d}{d\eta}+k^2{c_s}^2\right)(\Theta_0+\phi)=G(\eta)$$
$$G(\eta)\equiv\frac{k^2}{3}\left[\frac{1}{1+R}\phi-\psi\right]=k^2{c_s}^2\left[\phi-\psi(1+R)\right], \tag{7.47}$$

when ϕ,ψ are not constant but varying with time by adding the particular solution obtained by variation of parameters to the solution of the homogenous equation with $F=0$. Denoting the solutions of homogeneous equation by $S(\eta)\equiv\sin(kr_s)$ and $C(\eta)\equiv\cos(kr_s)$ The particular solution is given by

$$\Theta_0(\eta)+\phi(\eta)|_p=u(\eta)C(\eta)+v(\eta)S(\eta), \tag{7.48}$$

where u,v are the solutions of system of equations

$$u'(\eta)C(\eta)+v'(\eta)S(\eta)=0$$
$$u'(\eta)C'(\eta)+v'(\eta)S'(\eta)=G(\eta) \tag{7.49}$$

We can formally solve the above system to get

$$u'=\frac{G(\eta)S(\eta)}{C'(\eta)S(\eta)-S'(\eta)C(\eta)}=\frac{-G(\eta)S(\eta)}{kc_s}$$
$$v'=\frac{G(\eta)C(\eta)}{C'(\eta)S(\eta)-S'(\eta)C(\eta)}=\frac{G(\eta)C(\eta)}{kc_s}$$
$$u(\eta)=\int_0^\eta d\bar\eta\frac{-G(\bar\eta)S(\bar\eta)}{kc_s}$$
$$v(\eta)=\int_0^\eta d\bar\eta\frac{G(\bar\eta)C(\bar\eta)}{kc_s}. \tag{7.50}$$

We therefore get the general solution, keeping only the cosine mode of homogeneous solution as required by the adiabatic initial conditions,

$$\Theta_0(\eta)+\phi(\eta)=\left[\Theta_0(0)+\phi(0)\right]\cos(kr_s)+\int_0^\eta d\bar\eta\frac{G(\bar\eta)}{kc_s}\left[C(\bar\eta)S(\eta)-S(\bar\eta)C(\eta)\right]$$
$$=\left[\Theta_0(0)+\phi(0)\right]\cos(kr_s)$$
$$+k\int_0^\eta d\bar\eta\, c_s(\bar\eta)\left[\phi(\bar\eta)-\psi(\bar\eta)(1+R(\bar\eta))\right]\sin\left[k\left(r_s(\eta)-r_s(\bar\eta)\right)\right] \tag{7.51}$$

This is the general solution and in appropriate limits it reduces to the solutions derived in the previous sections.

We see that when the gravitational potentials ϕ, ψ are not constant, the particular solution will add a term that will in general out of phase with the first term coming from the homogeneous solution and the acoustic oscillations will acquire a phase shift. One particular case where we must take into account the changing potentials is for the contribution of neutrinos to ϕ, ψ which we have ignored so far. On superhorizon scales, initial perturbations in neutrinos as same as that in photons, due to adiabatic initial conditions, and therefore they also contribute to the metric perturbations ϕ, ψ. However, once neutrinos have decoupled, at $T \lesssim 1$ MeV, they are free streaming with the speed of light $c = 1 > c_{\rm s} \leq 1/\sqrt{3}$. Therefore, as soon as a mode enter horizon, i.e. at $\eta > 1/k$, the initial perturbations in neutrinos on scale k are erased at the speed of light. Thus the neutrino contribution to ϕ, ψ decays at the speed of light which is much faster than the speed of sound in the plasma. From Eq. 7.51 we see that if ϕ, ψ go from a finite value to zero in a short time, the integrand is non-zero only for part of the integration interval and after integration we will have a cosine as well as a sine contribution, resulting in a phase shift w.r.t the pure cosine mode [58].

7.1.6 Solution at second order in $R/\tau'_{\rm T}$ - Silk damping

So far we have studied the tightly coupled baryon-photon fluid in the ideal hydrodynamic limit. In particular we have ignored dissipative processes such as shear viscosity and thermal conductivity. We will now relax these assumptions. However, first we need a general form of the Boltzmann equation, which includes the energy and momentum conservations equations, Eqs. 7.18 and 7.19 as a special case. For dark matter as well as photons, we obtained the relativistic versions of the continuity and Euler equations by taking moments of the Boltzmann equation, Eq. 6.21 with the left hand side given by 6.34. For the photons, however, we do not need to integrate over the momentum. Substituting relation Eq. 7.6 in the left hand side of Boltzmann equation. Eq. 6.34, with $E = p$ and $\hat{p}^i\partial/\partial x^i \rightarrow ik\mu$ in Fourier space we get an equation for the evolution of temperature perturbations, $\Theta(\mathbf{k}, t, \mu)$, keeping only the first order terms, and including the collision term [see 47, for derivation] on the right hand side,

$$\Theta' + ik\mu\Theta + \phi' + ik\mu\psi = -\tau'_{\rm T}\left[\Theta_0 - \Theta + \mu v_{\rm b} - \frac{1}{2}P_2(\mu)\Theta_2\right]. \tag{7.52}$$

Note that once we substitute Eq. 7.6 in the Boltzmann equation, all first order terms are proportional to $p\partial f_0/\partial p$, where f_0 is the zeroth order Planck spectrum. It turns out that the collision term on the right hand side is also $\propto p\partial f_0/\partial p$, so that this factor cancels out throughout the equation and we do not need to integrate over the momentum p. Note that the collision term in the brackets on the right hand side is gauge invariant. This can be seen by substituting the Legendre polynomial decomposition for Θ, the Θ_0 term cancels out and combination of $\upsilon_{\rm b}$ and Θ_1 is gauge invariant as are all $\ell \geq 2$ modes. The term $\propto P_2$ arises from the angular dependence of the Thomson cross section $\propto 1 + \cos^2(\hat{\mathbf{p}}.\hat{\mathbf{p}}')$, where $\hat{\mathbf{p}}$ and $\hat{\mathbf{p}}'$ are the directions of photon before and after Thomson scattering.

Before going further, lets try to understand the collision term in Eq. 7.52 and the effect of Thomson scattering on the CMB. In order to do so, lets substitute the Legendre polynomial expansion for $\Theta(k, \mu, \eta)$ in Eq. 7.52. The Θ_0 term cancels and the dipole term appears in combination with the baryon velocity term.

$$\begin{aligned}\Theta' + ik\mu\Theta + \phi' + ik\mu\psi &= -\tau_{\rm T}'\left[\Theta_0 - \sum_{\ell}(-i)^{\ell}(2\ell+1)\Theta_{\ell}(k,\eta)P_{\ell}(\mu) + \mu\upsilon_{\rm b} - \frac{1}{2}P_2(\mu)\Theta_2\right]\\ &= n_{\rm e}\sigma_{\rm T}a\left[-\sum_{\ell=3}^{\infty}(-i)^{\ell}(2\ell+1)\Theta_{\ell}(k,\eta)P_{\ell}(\mu)\right.\\ &\left.-3(-i)\left(\Theta_1 - \frac{i\upsilon_{\rm b}}{3}\right)P_1(\mu) - 5(-i)^2\frac{9}{10}\Theta_2 P_2(\mu)\right] \qquad (7.53)\end{aligned}$$

First, note that if the radiation field is isotropic in the electron rest frame, i.e. $\Theta_\ell = 0$ for $\ell \geq 2$ and the CMB dipole in the electron rest frame vanishes, $\Theta_1 = i\upsilon_{\rm b}/3$ then we see that the collision term, right hand side of Eq. 7.52 or Eq. 7.53, also vanishes. Thus, Thomson scattering has no effect on the isotropic radiation field. This is due to the fact that the Thomson cross section depends on the relative direction (and not absolute direction) between the incoming and outgoing photons in the *electron rest frame*, and is symmeteric w.r.t the directions. Thus the cross section for a photon from direction $\hat{\mathbf{p}}$ to scatter into direction $\hat{\mathbf{p}}'$ is same as the cross section for a photon coming from direction $\hat{\mathbf{p}}'$ to scatter into direction $\hat{\mathbf{p}}$. If in addition, the radiation field is also isotropic, then the number of photons coming from the two directions are also equal. Therefore the number of photons scattered by the electron from direction $\hat{\mathbf{p}}$ into $\hat{\mathbf{p}}'$ is balanced by the photons scattered from $\hat{\mathbf{p}}'$ to $\hat{\mathbf{p}}$, resulting in no net change in the radiation field. This is of course true from all directions $\hat{\mathbf{p}}$ and $\hat{\mathbf{p}}'$. Thus an isotropic radiation field in the *electron rest frame* is unaffected by Thomson scattering and this results in the vanishing of the collision term in the Boltzmann equation. We also note again that Θ_1 is the photon dipole in the rest frame defined by the conformal Newtonian gauge metric while $\Theta_1 - i\upsilon_{\rm b}/3$ is the CMB dipole seen by the electrons/baryons with velocity $\upsilon_{\rm b}$. Note that unlike thermal motion where different particles would have equal energy at same temperature and therefore lighter particles would be moving faster, for bulk or average motion of the baryon-electron plasma, the velocities of all particles are equal irrespective of the mass. Thus electrons and ions have the same peculiar velocity v_b.

If the radiation field seen by the electrons is not isotropic, then the Boltzmann equation, ignoring the metric perturbations, will give $d\Theta/d\eta \propto -\Theta$. This is the decay equation and Thomson scattering will therefore result in the anisotropies to decay exponentially. Thus the Thomson scattering will attempt to isotropize the radiation field in the electron rest frame and drive all $\ell > 0$ modes in the electron rest frame to zero. Again, we can understand this very simply by looking at two directions $\hat{\mathbf{p}}$ and $\hat{\mathbf{p}}'$ which have different number of photons as seen by the electron with say more photons coming from the direction $\hat{\mathbf{p}}$ compared to $\hat{\mathbf{p}}'$. Since the cross section is symmetric w.r.t the two directions, we will have more photons scattering from $\hat{\mathbf{p}}$ into $\hat{\mathbf{p}}'$ than vice versa. Thus the effect of the Thomson scattering would be to reduce the number of photons in momentum directions which have higher number of photons and increase the number of photons in momentum directions which have a smaller number, isotropizing the radiation field. The equlibrium solution for Thomson scattering is an isotropic radiation field and this is the solution towards which evolution takes the photon-baryon system. We will see below explicitly, that when Thomson scattering is faster compared to the expansion rate, $\ell \geq 2$ multipoles of the CMB are suppressed. We have already seen that the leading contribution to the dipole is from the acoustic motions of the fluid. Finally the $\mu k\Theta$ term on the right hand side couples the neighbouring ℓ modes, as you will see in the following exercise.

Multiplying Eq. 7.52 by Legendre polynomial $P_\ell(\mu)$ and integrating over μ we get the Boltzmann hierarchy.

Exercise 26

Obtain the Boltzmann hierarchy from Eq. 7.52 by multiplying it by Legendre polynomial $P_\ell(\mu)$ and integrating over μ. Show that for $\ell = 0$ we recover Eq. 7.18. For $\ell = 1$ we obtain Eq. 7.19 with an extra term $\propto \Theta_2$ that we had ignored so far,

$$\Theta_1' + \frac{k}{3}(2\Theta_2 - \Theta_0) - \frac{k\psi}{3} = \tau_T'\left(\Theta_1 - \frac{i v_b}{3}\right) \tag{7.54}$$

and for $\ell \geq 2$ we have

$$\Theta_2' + \frac{k}{5}(3\Theta_3 - 2\Theta_1) = \frac{9}{10}\tau_T'\Theta_2 \tag{7.55}$$

$$\Theta_\ell' + \frac{k}{2\ell+1}\left[(\ell+1)\Theta_{\ell+1} - \ell\Theta_{\ell-1}\right] = \tau_T'\Theta_\ell \quad : \ell \geq 3 \tag{7.56}$$

Hint: The first two Legendre polynomials are

$$P_0(\mu) = 1,\; P_1(\mu) = \mu \tag{7.57}$$

The orthogonality relation for Legendre polynomials is

$$\int_{-1}^{1} \mathrm{d}\mu P_\ell(\mu) P_{\ell'}(\mu) = \delta_{\ell\ell'} \frac{2}{2\ell+1} \tag{7.58}$$

and they satisfy the recurrence relation

$$(\ell+1)P_{\ell+1}(\mu) = (2\ell+1)\mu P_\ell(\mu) - \ell P_{\ell-1}(\mu). \tag{7.59}$$

Formally, the Boltzmann hierarchy is an infinte series in ℓ with each ℓ mode coupled to $\ell-1$ and $\ell+1$ mode. In order to solve it we must truncate the hierarchy at some finite ℓ. So far we have been truncating the hierarchy at $\ell = 1$, assuming that $\Theta_\ell = 0$ for $\ell \geq 2$. Lets now include the $\ell = 2$ mode. Note that the assumption $\Theta_\ell \approx 0$ is a good assumption in the tight coupling limit, since when $\tau'_{\mathrm{T}}\eta \gg 1$, $\Theta_\ell \to 0$ so that the right hand side in Eq. 7.56 remains finite. In particular Θ_ℓ is suppressed more and more with increasing ℓ by powers of τ'_{T}. In fact we can see from Eq. 7.56 that with $\Theta'_\ell \sim \Theta_\ell/\eta$,

$$\Theta_\ell \sim \frac{\ell}{2\ell+1}\frac{k}{\tau'_{\mathrm{T}}}\Theta_{\ell-1}. \tag{7.60}$$

Thus, including higher ℓ modes is equivalent to going to higher order in k/τ'_{T}. Note that τ'_{T} is negative and therefore the Thomson collision term acts so as to reduce the amplitude of Θ_ℓ for $\ell \geq 2$.

The metric perturbations do not play an important role in dissipation and we can ignore them in this subsection. Ignoring the metric perturbations and $\ell > 2$ modes, we have the following system of equations to solve:

$$\Theta'_0 + k\Theta_1 = 0 \tag{7.61}$$

$$\Theta'_1 + \frac{k}{3}(2\Theta_2 - \Theta_0) = \tau'_{\mathrm{T}}\left(\Theta_1 - \frac{i\upsilon_{\mathrm{b}}}{3}\right) \tag{7.62}$$

$$\Theta'_2 - \frac{2k}{5}\Theta_1 = \frac{9}{10}\tau'_{\mathrm{T}}\Theta_2. \tag{7.63}$$

As before, we want to expand $\upsilon_{\mathrm{b}} + 3i\Theta_1$ in powers of $1/\tau'_{\mathrm{T}}$ bu going to $(1/\tau'_{\mathrm{T}})^2$ order now. The equation for υ_{b} is

$$\upsilon_{\mathrm{b}} + 3i\Theta_1 = \frac{R}{\tau'_{\mathrm{T}}}\left[\upsilon_{\mathrm{b}}{}' + \frac{a'}{a}\upsilon_{\mathrm{b}}\right] \tag{7.64}$$

We want solutions of the form $e^{i\int \omega \mathrm{d}\eta}$, where the real part of $\omega \approx kc_{\mathrm{s}}$ as before is the oscillatory part of the solution and imgainary part gives dissipation (or exponential growth). Microscopically, damping happens because of photons diffusing through the electrons on scales much smaller compared to the horizon size or at high frequencies so that $v_{\mathrm{b}} = i\omega v_{\mathrm{b}} \gg v_{\mathrm{b}}a'/a$ and we can ignore the Hubble damping term when studying dissipation. In fact photons are doing a random walk through the plasma with step size of order mean free path, λ_{mfp} and in N steps they can diffuse through a distance of order $\sim \sqrt{N}\lambda_{\mathrm{mfp}}$ and number of steps photons can take in time η is $\sim \eta/\lambda_{\mathrm{mfp}}$ giving the diffusion distance to be $\sim \sqrt{\eta\lambda_{\mathrm{mfp}}} \sim \sqrt{\eta/(-\tau'_{\mathrm{T}})}$ (remember $\tau'_{\mathrm{T}} = -n_{\mathrm{e}}\sigma_{\mathrm{T}}a$, Eq. 7.15).

Ignoring the damping term and substituting solutions $v_{\mathrm{b}}, \Theta_1 \propto e^{i\int \omega \mathrm{d}\eta}$ in Eq. 7.64 gives

$$v_{\mathrm{b}}\left(1 - \frac{i\omega R}{\tau'_{\mathrm{T}}}\right) = -3i\Theta_1 \tag{7.65}$$

Solving for v_{b} and expanding in powers of $\omega R/\tau'_{\mathrm{T}}$ we get

$$\begin{aligned} v_{\mathrm{b}} &= -3i\Theta_1\left(1 - \frac{i\omega R}{\tau'_{\mathrm{T}}}\right)^{-1} \\ &\approx -3i\Theta_1\left[1 + \frac{i\omega R}{\tau'_{\mathrm{T}}} - \left(\frac{\omega R}{\tau'_{\mathrm{T}}}\right)^2\right] \\ \frac{iv_{\mathrm{b}}}{3} - \Theta_1 &= \Theta_1\left[\frac{i\omega R}{\tau'_{\mathrm{T}}} - \left(\frac{\omega R}{\tau'_{\mathrm{T}}}\right)^2\right] \end{aligned} \tag{7.66}$$

Also $\Theta'_2 \ll \tau'_{\mathrm{T}}\Theta_2$ for τ'_{T} large. Ignoring Θ'_2 in Eq. 7.55 we get solution for Θ_2,

$$\Theta_2 = -\frac{4k}{9\tau'_{\mathrm{T}}}\Theta_1. \tag{7.67}$$

Also Eq. 7.61 gives on substituting solutions $\propto e^{i\int \omega \mathrm{d}\eta}$ for all variables

$$i\omega\Theta_0 = -k\Theta_1 \tag{7.68}$$

and using these solutions in Eq. 7.62 gives us the dispersion relation

$$i\omega + \frac{k}{3}\left[2\left(\frac{-4}{9}\right) + \frac{k}{i\omega}\right] = -\tau'_{\mathrm{T}}\left[\frac{i\omega R}{\tau'_{\mathrm{T}}} - \left(\frac{\omega R}{\tau'_{\mathrm{T}}}\right)^2\right] \tag{7.69}$$

Rearranging the terms, the above equation can be written as

$$\omega^2 (1+R) - \frac{k^2}{3} + \frac{i\omega}{\tau'_{\rm T}}\left(\omega^2 R^2 + \frac{8k^2}{27}\right) = 0 \tag{7.70}$$

The zeroth order term in $1/\tau'_{\rm T}$ just gives us $\omega^2 = k^2 {c_{\rm s}}^2 \equiv \omega_0^2$. Writing $\omega = \omega_0 + \delta\omega$, where $\delta\omega$ is a small correction to ω_0, we get at first order in $\delta\omega$ and $1/\tau'_{\rm T}$,

$$2\omega_0\delta\omega(1+R) = \frac{-i\omega_0}{\tau'_{\rm T}}\left(\omega_0^2 R^2 + \frac{8k^2}{27}\right) \tag{7.71}$$

Rearranging terms, we can write the solution for $\delta\omega$ as

$$\delta\omega = \frac{ik^2}{6n_{\rm e}\sigma_{\rm T}a(1+R)}\left(\frac{R^2}{1+R} + \frac{8}{9}\right) \tag{7.72}$$

We therefore have the solutions for $\Theta_0, \Theta_1 v_{\rm b}$ of the form

$$\begin{aligned} &\propto e^{i\int \omega {\rm d}\eta} = e^{ikr_{\rm s}} e^{\int \delta\omega {\rm d}\eta} \\ &= e^{ikr_{\rm s}} e^{-k^2/{k_{\rm D}}^2}, \end{aligned} \tag{7.73}$$

where $1/k_{\rm D}$ is given by

$$\frac{1}{{k_{\rm D}}^2} = \int_0^\eta {\rm d}\bar{\eta} \frac{1}{6n_{\rm e}\sigma_{\rm T}a(1+R)}\left(\frac{R^2}{1+R} + \frac{8}{9}\right) \tag{7.74}$$

and the diffusion length is given by $\lambda_{\rm D} = 2\pi/k_{\rm D}$. This exponential damping of the sound waves on small scales, for $k > k_{\rm D}$ or $\lambda < \lambda_{\rm D}$ is known as Silk damping [59, 60]. Thomson scattering of initially unpolarized photons creates polarization if there is non-zero quadrupole in the radiation field. Therefore when we include quadrupole, we should also take into account polarization. Including polarization effects changes the factor of 8/9 in the expression for $k_{\rm D}$ to 16/15 [61]. If we assume $R \approx 0$ and fully ionized plasma, Eq. 7.74 can be integrated analytically.

Exercise 27

Integrate Eq. 7.74 assuming $R \approx 0$ and fully ionized plasma. What are the values of $k_{\rm D}$ and $\lambda_{\rm D}$ at the time of recombination, $z \approx 1000$. Use values of cosmological parameters from Planck 2018 cosmological parameters paper for numerical results and give results in units of Mpc.

7.2 Mixing of blackbodies and Sunyaev-Zeldovich effect

So far we have assumed that the CMB spectrum is a Planck spectrum at every point in space **x**. This is an excellent approximation, since in the early Universe every region was in local thermal equilbrium and a Planck spectrum was established for the photons. However, since the energy density in different parts of the Universe had fluctuations, the temperature of the blackbody spectrum of the CMB was also slightly different in different parts of the Universe. Since the CMB has a blackbody spectrum we can talk about temperature fluctuations instead of energy density fluctuations and in particular we do not need to keep track of the full spectrum, $f(p)$.

This is all very well until we come to the Silk damping or dissipation of sound waves due to photon diffusion. We saw in the previous section that on diffusion scales, $\lambda_{\rm D} = 2\pi/k_{\rm D}$ the photons from different regions having different temperature mix together resulting in erasure of CMB perturbations on these scales. Mixing of CMB photons on scales $\lesssim \lambda_{\rm D}$ implies that CMB on these scales would become homogeneous with the same energy density and spectrum everywhere. However, the spectrum of these photons, after averaging different temperature blackbodies, will not be a blackbody spectrum. We can see this by considering averaging of just two blackbodies with temperature T_1 and T_2. The energy density of photons of a blackbody with temperature T is $a_{\rm R}T^4$, where $a_{\rm R} = 4\sigma_{\rm SB}/c = \frac{8\pi^5 {k_{\rm B}}^4}{15c^3h^3}$ is the radiation constant, $\sigma_{\rm SB}$ is the Stefan-Boltzmann constant. The number density of photons is $b_{\rm R}T^3$, where $b_{\rm R} = \frac{16\pi {k_{\rm B}}^2\zeta(3)}{c^3h^3}$, ζ is the Riemann zeta function with $\zeta(3) = 1.20206$. After mixing, the energy density of photons is $(1/2)a_{\rm R}\left(T_1^4 + T_2^4\right)$ and the number density of photons is $(1/2)b_{\rm R}\left(T_1^3 + T_2^3\right)$. Suppose, the final spectrum after mixing was also blackbody spectrum with temperature T, then conservation of energy implies that $T = \left[1/2\left(T_1^4 + T_2^4\right)\right]^{1/4}$ while conservation of number of photons implies that $T' = \left[1/2\left(T_1^3 + T_2^3\right)\right]^{1/3} \neq T$. Thus, if both energy and photon number are conserved, the final spectrum cannot by blackbody. The only way a blackbody spectrum can be established is if photon number is also violated which can happen only at $z \gtrsim 2 \times 10^6$ when processes such as double Compton scattering and bremsstrahlung are efficient [62–65].

We can calculate the spectrum resulting from the mixing of blackbodies during photon diffusion before recombination by averaging the phase space density or the occupation number of photons. The photons mixing together belong to blackbody of temperature $T = \bar{T} + \Delta T = \bar{T}\,(1 + \Theta)$, where $\bar{T}$ is the average temperature and ΔT is the deviation from it which in standard cosmology has a Gaussian probability density function (PDF) and we must average the phase space density over this distribution of temperature. We will denote the average of temperature PDF by angular brackets. Thus the spectrum after mixing, $f(\bar{T})$ is given by

$$\bar{f}(T) = \langle f(T)\rangle = \left\langle \frac{1}{e^{\frac{h\nu}{k_{\rm B}(\bar{T}+\Delta T)}} - 1} \right\rangle. \tag{7.75}$$

We can treat the occupation number f as a function of $\ln T$ and write $f(T) \equiv f(\ln T) = f(\ln\bar{T} + \ln(1+\Theta))$. It is now easy to do a Taylor series expansion around $\ln\bar{T}$ giving,

$$\begin{aligned}\bar{f}(T) &= \left\langle f(\ln\bar{T} + \ln(1+\Theta))\right\rangle \\ &= \left\langle f(\bar{T}) + \ln(1+\Theta)\frac{\partial f(\bar{T})}{\partial\ln\bar{T}} + \frac{1}{2}\left[\ln(1+\Theta)\right]^2\frac{\partial^2 f(\bar{T})}{\partial\ln(\bar{T})^2} + \text{ higher order terms}\right\rangle.\end{aligned} \tag{7.76}$$

Expanding the $\ln(1+\Theta)$ also in Taylor series and keeping terms upto second order in Θ, we get after some rearrangement of terms,

$$\begin{aligned}\bar{f}(T) &= f(\ln\bar{T}) + \left(\langle\Theta\rangle + \langle\Theta^2\rangle\right)\bar{T}\frac{\partial f(\bar{T})}{\partial\bar{T}} + \frac{1}{2}\langle\Theta^2\rangle\bar{T}^4\frac{\partial}{\partial\bar{T}}\frac{1}{\bar{T}^2}\frac{\partial f(\bar{T})}{\partial\bar{T}} \\ &= f\left(\bar{T}\left(1+\langle\Theta^2\rangle\right)\right) + \frac{1}{2}\langle\Theta^2\rangle f_{\rm sz},\end{aligned} \tag{7.77}$$

where we have used that by definition $\langle\Theta\rangle = 0$, and the first two terms in the first line are equivalent to a Planck spectrum with temperature $\bar{T}\left(1+\langle\Theta^2\rangle\right)$ up to second order in Θ. We have also defined the Sunyaev-Zeldovich spectrum,

$$\begin{aligned}f_{\rm sz} &= \bar{T}^4\frac{\partial}{\partial\bar{T}}\frac{1}{\bar{T}^2}\frac{\partial f(\bar{T})}{\partial\bar{T}} \\ &= \frac{1}{x^2}\frac{\partial}{\partial x}x^4\frac{\partial f(x)}{\partial x},\end{aligned} \tag{7.78}$$

where in the last line we have done a change of variables from $\bar{T}$ to $x = \frac{h\nu}{k_{\rm B}\bar{T}}$ so that $\partial/\partial\bar{T} = -(x/\bar{T})\partial/\partial x$. For Planck spectrum $f(x) = 1/(e^x - 1)$ and

$$f_{\rm sz} = \frac{xe^x}{(e^x-1)^2}\left[x\left(\frac{e^x+1}{e^x-1}\right) - 4\right]. \tag{7.79}$$

We have derived the expression for the Sunyaev-Zeldovich (SZ) spectral distortion in Eq. 7.79. The amplitude of the distortion, y is give by half the variance of the temperature anisotropies $y = \langle\Theta^2\rangle/2$. The definition of the Sunyaev-Zeldovich spectrum originates in the Compton scattering of CMB photons with a hotter plasma. In such an interaction, electrons will add energy to the CMB photons, boosting them. The number of photons however does not change in Compton scattering. The Sunyaev-Zeldovich distortion defined in Eq. 7.79 is the difference between the final spectrum and the initial blackbody spectrum. Since both the initial and final spectrum have the same number density of photons, the number density of photons in their difference should be zero. The number density of photons is $\propto \int \mathrm{d}x x^2 f(x)$, for any spectrum $f(x)$. For the SZ spectrum, the number density, $N_{\rm SZ}$ is therefore

$$\begin{aligned} N_{\rm SZ} &\propto \int_0^\infty \mathrm{d}x x^2 f_{\rm sz}(x) \\ &\propto \int_0^\infty \mathrm{d}x \frac{\partial}{\partial x} x^4 \frac{\partial f(x)}{\partial x} \\ &\propto x^4 \frac{\partial f(x)}{\partial x}\bigg|_0^\infty = 0 \end{aligned} \tag{7.80}$$

The energy density in a spectrum with occupation number $f(x)$ is propto $\int \mathrm{d}x x^3 f(x)$. The energy density in the SZ distortion, $E_{\rm SZ}$, is

$$\begin{aligned} E_{\rm SZ} &\propto \int_0^\infty \mathrm{d}x x^3 f_{\rm sz}(x) \\ &= \int_0^\infty \mathrm{d}x x \frac{\partial}{\partial x} x^4 \frac{\partial f(x)}{\partial x} \\ &= \int_0^\infty \mathrm{d}x x^4 \frac{\partial f(x)}{\partial x} \\ &= \int_0^\infty \mathrm{d}x 4x^3 f(x) \\ &\propto 4E_{\rm BB}(\bar{T}), \end{aligned} \tag{7.81}$$

where $E_{\rm BB}$ is the energy density of the blackbody spectrum. We have integrated by parts and used that the boundary terms of the type $x^4 f(x)$, $x^5 f(x)$ vanish at $x = 0, \infty$ when $f(x)$ is the Planck spectrum. Thus the fractional energy gained by the CMB due to the SZ distortion is, $\Delta E/E = E_{\rm SZ}/E_{\rm BB} = 4y = 2\langle\Theta^2\rangle$. Note from Eq. 7.77 that part of the dissipation energy goes into raising the the temperature of the blackbody part of the CMB by $\langle\bar{T}\Theta^2\rangle$. The absolute value of the energy density is a frame dependent quantity and will be different in a different gauge. However, the relative change in the energy density, and therefore the amplitude y of the SZ distortion is gauge invariant. The Sunyaev-Zeldovich effect is also referred to as the y-type distortion in literature because the original papers of Sunyaev and Zeldovich [66] as well as subsequent work on the subject used the symbol y to refer to the amplitude of the distortion. The CMB anisotropies, Θ are of order 10^{-4} and therefore the amplitude of distortion we expect from Silk damping of CMB anisotropies would be of order $y \sim \Theta^2 \sim 10^{-8}$. Note that we have statistical isotropy in the Universe which means that ampliitude of the small scale anisotropies of the CMB should be same in all parts of the Universe. The amplitude of the CMB spectral distortion should also be isotropic at lowest order i.e. this distortion would be same in all directions and would be observable as the deviation of the average CMB or CMB monopole from a blackbody. The deviation of the CMB monopole from a blackbody has not yet been detected and the current best constraints come from Far Infrared Absolute Spectrophotometer (FIRAS) ecperiment onboard the first CMB space mission, the Cosmic Background Explorer (COBE) satellite and imply $y \lesssim 10^{-5}$. We therefore need a few orders of magnitude improvement in the sensitivity in order to be able to detect the spectral distortions from Silk damping.

7.2.1 Kompaneets equation

In the standard ΛCDM cosmological model, even though the early Universe is in almost perfect equilibrium and the spectrum of the CMB is almost perfect blackbody, subsequent evolution and interaction of the CMB with matter can create deviations from the blackbody spectrum. These deviations are often referred to as the spectral distortions of the CMB. The Sunyaev-Zeldovich effect is an example of such a spectral distortion, in fact the only type of spectral distortion that has been detected so far, although other type of deviations are possible and in fact must exist also. In particular, the Sunyaev-Zeldovich effect has been detected in the direction of clusters of galaxies, where the interaction of the CMB with the hot gas in the intergalactic (or intracluster) medium creates a y-type distortion.

A more direct and traditional way to understand the Sunyaev-Zeldovich effect from the clusters of galaxies is using the Boltzmann equation rather than mixing of blackbodies. We saw above that the spectral distortion appears at second order the temperature perturbations. We can ignore the metric perturbations as well as the CMB anisotropies i.e. assume the CMB is isotropic. The Boltzmann equation is then given by

$$\frac{\partial f}{\partial t} = C(f) \tag{7.82}$$

As we saw earlier, fif the CMB is isotropic, the collision term vanished at linear order. We must therefore go to the second order in perturbations in order get a non-zero contribution from the collision term. This is consistent with our expectation from mixing of blackbodies. Alternatively, since we are interested in energy transfer between photons and electrons, and in the non-relativistic limit the energy transfer is of order p/m_e or T_e/m_e, where p is the energy of CMB photons, T_e is the electron temperature. We must therefore calculate the collision term to at first order in p/m_e and T_e/m_e. The interaction of photons and electrons results in change in energy of photon in each scattering. In the limit of large number of scatterings such that in each scattering the photon energy changes by a small amount, $\Delta p/p \ll 1$. The photons therefore do a random walk in the momentum space, gaining or loosing a small amount of energy in each scattering. In the language of random walks, we want to do a Fokker-Planck expansion of the Boltzmann equation, and calculate the collision integral upto second order in $\Delta p/p$. This calculation was first done by Kompaneets [67] and the resulting Fokker-Planck equation is called Kompaneets equation.

$$\frac{\partial f}{\partial t} = n_e \sigma_T c \frac{k_B T}{m_e c^2} \frac{1}{x^2} \frac{\partial}{\partial x} x^4 \left(f + f^2 + \frac{T_e}{T} \frac{\partial f}{\partial x} \right), \tag{7.83}$$

where T is the CMB temperature and T_e is the electrons temperature. The first term in the brackets on the right side is recoil term due to transfer of energy from photon to electron in the collision, the second term is stimulated recoil term due to the presence of photons of the same energy as the outgoing photon in the radiation field. The T_e/T term is due to the energy transfer from electrons to photons and vanished if the electrons are at rest, i.e. $T_e = 0$. The energy transfer is due to the Doppler effect from thermal motion of electrons and it is of second order in electron thermal velocities, $T_e \propto v^2$.

Exercise 28

Show that the Bose-Einstein spectrum is an equilibrium solution of the Kompaneets equation.

In clusters of galaxies, $T_e \gg T$, and we can ignore the recoil terms compared to the Doppler term. This reduces the Kompaneets equation to a diffusion equation.

$$\frac{\partial f}{\partial t} = n_e \sigma_T c \frac{k_B T_e}{m_e c^2} \frac{1}{x^2} \frac{\partial}{\partial x} x^4 \left(\frac{\partial f}{\partial x} \right) \tag{7.84}$$

The above equation can be converted to standard heat equation by a change of variables to which a formal solution exists [62]. We can find a simpler solution in the limit when $y = n_e \sigma_T c k_B T_e/(m_e c^2) \Delta t \ll 1$, where Δt is the light crossing time of cluster. In this limit, the distortion of the CMB will be small and we can solve the equation by just substituting the initial Planck spectrum on the right hand side of Eq. 7.84. This is equivalent to taking a signle numerical time step in solving the partial differential equation. We therefore get,

$$\begin{aligned}\frac{\partial f}{\partial t} &= n_{\rm e}\sigma_{\rm T}c\frac{k_{\rm B}T_{\rm e}}{m_{\rm e}c^2}\frac{1}{x^2}\frac{\partial}{\partial x}x^4\left(\frac{\partial f_{\rm pl}}{\partial x}\right)\\ &= n_{\rm e}\sigma_{\rm T}c\frac{k_{\rm B}T_{\rm e}}{m_{\rm e}c^2}f_{\rm SZ}\end{aligned} \tag{7.85}$$

which is readily integrated to give,

$$\begin{aligned}\Delta f &= \int {\rm d}t\, n_{\rm e}\sigma_{\rm T}c\frac{k_{\rm B}T_{\rm e}}{m_{\rm e}c^2}f_{\rm SZ}\\ &= y_{\rm SZ}f_{\rm SZ}.\end{aligned} \tag{7.86}$$

The amplitude of the distortion is

$$y_{\rm SZ} = \int {\rm d}t\, n_{\rm e}\sigma_{\rm T}c\frac{k_{\rm B}T_{\rm e}}{m_{\rm e}c^2}, \tag{7.87}$$

where the integral is over the cluster or as long as CMB is interacting with hot electrons.

The fact that we have recovered the same formula for the spectral distortion that we have obtained from mixing of blackbodies, with the amplitude $\propto T_{\rm e} \propto \upsilon^2$, is not a coincidence. We can think of interaction of the isotropic CMB radiation with hot electrons also as mixing of blackbodies. If we go to the reference frame of any electron, it will see a Doppler shifted CMB, with higher than average temperature in the direction it was moving and lower temperature in the other direction. The *Thomson* scattering of this dipolar radiation field with the electron will result in mixing of blackbodies. As we argued above, the shape of the CMB distortion and its dimensionless amplitude does not depend on the frame of reference. If the electrons have isotropic thermal distribution, the CMB dipole anisotropy seen by the electrons will be proportional to the electron velocity υ which will average to zero at linear order (isotropy) but at second order the average will be $\propto \langle\upsilon^2\rangle \propto T_{\rm e}$. Note that the amplitude of the y-type distortion from a cluster is proportional to the integral of the pressure $P_{\rm e} = n_{\rm e}k_{\rm B}T_{\rm e}$ along the line of sight through the cluster.

7.3 CMB anisotropies on Super horizon scales - Sachs-Wolfe effect

So far we have studied the CMB anisotropies on sub-horizon scales. We saw that the perturbations in baryon-photon plasma oscillate and would give the characteristic pattern of peaks and troughs in the CMB anisotropies. We also saw that the CMB anisotropies decay because of photon diffusion, causing the CMB anisotropies to decay exponentially on scales smaller than the diffusion scales.

To complete the discussion we will now look at the CMB anisotropies on scales large compared to the horizon size at recombination, i.e. scales k such that $k\eta_* \ll 1$, where η_* is the conformal time or comoving horizon at recombination, $z_* \approx 1100$. On superhorizon scales, interactions do not matter since they cannot influence scales out of causal contact but whether a particle is relativistic or non-relativistic matters. Thus the evolution of both radiation components, photons and neutrinos, is identical as is the evolution of baryons and dark matter. We will need the continuity equations for radiation Eq. 7.18 (with temperature perturbation Θ_0 same for neutrinos and photons) and matter with density and velocity perturbations δ and υ, Eq: 6.47

$$\Theta_0' + k\Theta_1 = -\phi' \tag{7.88}$$

$$\delta' + ik\upsilon = -3\phi' \tag{7.89}$$

In addition we will also need the time-time component of the Einstein's equations

$$k^2\phi + 3\frac{a'}{a}\left(\phi' - \psi\frac{a'}{a}\right) = 4\pi G a^2\left[\rho_{\rm m}\delta + 4\rho_{\rm r}\Theta_0\right], \tag{7.90}$$

where $\rho_{\rm m} = \rho_{\rm b} + \rho_{\rm cdm}$ is the total matter energy density and includes baryons and dark matter and $\rho_{\rm r} = \rho_\gamma + \rho_\nu$ is the total radiation energy density and includes photons and neutrinos. Note that on sub-horizon scales, $k\eta \gg 1$, or if we can ignore expansion, we can ignore the metric terms in the brackets on the right hand side compared to the $k^2\phi$ term and the above equation reduces to the Poisson equation or Newtonian gravity (after taking into account that the energy density in radiation also gravitates). On super horizon scales on the other hand, $k\eta \ll 1$, we can ignore the $k^2\phi$ term compared to the Hubble terms and the time-time component of the Einstein's equations becomes

$$3\frac{a'}{a}\left(\phi' - \psi\frac{a'}{a}\right) = 4\pi G a^2\left[\rho_{\rm m}\delta + 4\rho_{\rm r}\Theta_0\right] \tag{7.91}$$

In the coninuity equations, Eq: 7.89, also we can ignore the velocity terms, $k\upsilon$, $k\Theta_1$, on super horizon scales giving,

$$\Theta_0' = -\phi' \tag{7.92}$$

$$\delta' = -3\phi' \tag{7.93}$$

We are in particular interested in how the perturbations on superhorizon scales evolve as we go from radiation dominated era to matter dominated era when the recombination happens and CMB photons free stream to us. It is convenient to define the redshift or scale factor corresponding to the matter-radiation equality, $a_{\rm eq}$, when $\rho_{\rm m} = \rho_{\rm r}$. We can now define a new *time* variable, y,

$$y \equiv \frac{a}{a_{\rm eq}} = \frac{\rho_{\rm m}}{\rho_{\rm r}} \tag{7.94}$$

and use it instead of η to study the evolution of perturbations. The limit $y \ll 1$ corresponds to radiation domination and $y \gg 1$ corresponds to matter domination. The variable y is just the scale factor normalized to unity at matter-radiation equality instead of today. We have therefore just changed the normalization of the scale factor. The Einstein equation can now be written as, with the approximation $\phi = -\psi$,

$$\begin{aligned} 3\frac{a'}{a}\left(\phi' + \phi\frac{a'}{a}\right) &= 4\pi G a^2 \rho_{\rm m}\delta\left[1 + \frac{4}{3y}\right] \\ &= 4\pi G a^2\left(\rho_{\rm m} + \rho_{\rm r}\right)\frac{\rho_{\rm m}}{\rho_{\rm m} + \rho_{\rm r}}\delta\left[1 + \frac{4}{3y}\right] \\ &= \frac{3}{2}\left(\frac{a'}{a}\right)^2\frac{y}{1+y}\delta\left[1 + \frac{4}{3y}\right], \end{aligned} \tag{7.95}$$

where we have used the Friedmann equation in the last line. Changing variables from η to y, with

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\eta} &= \frac{\mathrm{d}y}{\mathrm{d}\eta}\frac{\mathrm{d}}{\mathrm{d}y} \\ &= y\frac{a'}{a}\frac{\mathrm{d}}{\mathrm{d}y}, \end{aligned} \tag{7.96}$$

we get for the Einstein equation

$$\begin{aligned} y\frac{\mathrm{d}\phi}{\mathrm{d}y} + \phi &= \frac{y}{2(1+y)}\delta\left[1 + \frac{4}{3y}\right] \\ &= \frac{3y+4}{6(y+1)}\delta \end{aligned} \tag{7.97}$$

Taking derivative w.r.t y and using $\mathrm{d}\delta/\mathrm{d}y = -3\mathrm{d}\phi/\mathrm{d}y$ from Eq. 7.93 we get,

$$\frac{\mathrm{d}}{\mathrm{d}y}\left[\frac{6(y+1)}{3y+4}\left(y\frac{\mathrm{d}\phi}{\mathrm{d}y} + \phi\right)\right] = \frac{\mathrm{d}\delta}{\mathrm{d}y} = -3\frac{\mathrm{d}\phi}{\mathrm{d}y} \tag{7.98}$$

Carrying out the differentiation and rearranging terms, we get a second order linear differential equation for ϕ,

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}y^2} + \frac{21y^2 + 54y + 32}{2y(y+1)(3y+4)}\frac{\mathrm{d}\phi}{\mathrm{d}y} + \frac{\phi}{y(y+1)(3y+4)} = 0 \tag{7.99}$$

It is possible to find an analytic solution to this complicated looking equation by doing a change of variables [46]

$$u = \frac{y^3}{\sqrt{1+y}}\phi \tag{7.100}$$

Exercise 29

Optional - Show that the solution to Eq. 7.99 is given by

$$\phi(y) = \frac{\phi(0)}{10}\frac{1}{y^3}\left[16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16\right] \tag{7.101}$$

Also show that as $y \to 0$, $\phi(y) \to \phi(0)$. Hint: See section 7.2.1 in Dodelson [47].

During matter domination, $y \gg 1$, and the $9y^3$ term in the barckets dominates giving $\phi(y \gg 1) \approx 9/10\phi(0)$. Thus the superhorizon metric perturbations decay a little bit as we go from radiation domination to matter domination. During recombination, $a \approx 10^{-3}$ and $y \approx 3$, and matter domination is a good approximation, $\phi(\eta_*) \approx 9/10\phi(0)$ for $k\eta_* \ll 1$.

We are now in position to derive the large scale solution first derived by Sachs and Wolfe [68] and known as the Sachs-Wolfe effect. Integrating Eq. 7.93, we get solution for Θ_0, $\Theta_0(\eta) = -\phi(\eta) +$ constant.

We have the adiabatic initial conditions, Eq. 7.32, at $\eta = 0$ with $\Theta_0(0) = \phi(0)/2$, therefore the constant is $3\phi(0)/2$,

$$\Theta_0(\eta) = -\phi(\eta) + \frac{3}{2}\phi(0) \tag{7.102}$$

The observed anisotropy at the last scattering surface, $\Theta_0(\eta_*) + \psi(\eta_*)$ is therefore given by (again using $\phi = -\psi$,

$$\begin{aligned}\Theta_0(\eta_*) + \psi(\eta_*) &= \Theta_0(\eta_*) - \phi(\eta_*)\\ &= -2\phi(\eta_*) + \frac{3}{2}\phi(0)\\ &= -\frac{1}{3}\phi(\eta_*) = \frac{1}{3}\psi(\eta_*).\end{aligned} \tag{7.103}$$

Thus in a region where there was an over density initially, so that $\phi(0) > 0$, $\psi(\eta_*) < 0$ and we will see a colder CMB. Similarly in an initial underdensity, we will see a higher temperature on superhorizon scales. This is opposite to what we see on sub-horizon scales, where initial perturbation fell into gravitational potential wells, and from our acoustic solution we see that initiai overdensities at the first acoustic peak would have higher temperature. On the superhorizon scale, the gravitational redshift or blueshift suffered by the CMB photons as they free stream towards us dominates over the initial over/under-density resulting in lower observed temperature where CMB had higher temperature at the last scattering surface and vice versa fpr lower temperature regions at the last scattering surface. Also the amplitude of the perturbations is only $(1/3)\psi$ instead of $(1/2)\psi$ that we would have expected just from the adiabatic initial conditions.

We should however be careful when using our Newtonian intuition to interpret the superhorizon modes [see 69, for an interesting discussion], in particular interpretation of the ψ term in the observable $\Theta_0 + \psi$ in terms of gravitational redshift suffered by the photons as they travel from the last scattering surface to us. However, we note that even if these modes our superhorizon at the time of recombination, they are subhorizon today, and it is today's subhorizon anisotropies measurable by an observer at our position that we are studying.

7.4 Line of sight solution to the CMB anisotropies

We have so far studied the evolution of the CMB perturbations starting with the initial conditions at the beginning of the radiation dominated phase of the Universe on superhorizon scales which were possibly the result of an even earlier inflationary phase of the Universe. We studied the evolution of the modes which entermhorizon before recombination and result in standing sound waves in the baryon photon plasma as well as superhorizon evolution of modes which are still outside the horizon at the time of recombination. We now have solutions for CMB monopole, dipole and also quadrupole at the time of recombination. However to connect with what we observe today, we must evolve these solutions from recombination until today, $\eta = \eta_0$.

We are thus interested in the CMB temperature field $\Theta(\mathbf{x}_0, \eta_0, \hat{\mathbf{n}})$, where $\mathbf{x}_0, \eta_0$ are the comoving space-time coordinates of the observer, i.e. us, with η_0 the conformal time today and $\hat{\mathbf{n}} \equiv (\theta, \phi)$ is the direction of observation where θ, ϕ are the angles of the spherical coordinate system centered at our position and we are observing the CMB photons coming from different directions $\hat{\mathbf{n}}$. Since the Universe is a random realization, with the CMB and other primordial fluctuations one particular realization of a (Gaussian) random field, we cannot of course predict CMB that will be observed at our position. What we can predict is the statistical properties of the underlying random field, such as its power spectrum or even test whether the random field is Gaussian or not. Decomposing the observed temperature anisotropies into spherical harmonics ($Y_{\ell m}$), we get,

$$a_{\ell m}(\mathbf{x}_0, \eta_0) = \int \mathrm{d}\hat{\mathbf{n}}\Theta(\mathbf{x}_0, \eta_0, \hat{\mathbf{n}})Y_{\ell m}(\hat{\mathbf{n}}). \tag{7.104}$$

Before we relate these observables, $a_{\ell m}$ to the transfer functions, Θ_ℓ that we have been calculating so far, we should note that we can observe modes upto $\ell \leq \ell_{\max}$, where $\ell_{\max}$ is determined by the angular resolution of the CMB experiments. The multipole number ℓ is the counterpart (on the surface of sphere) of the Fourier wavenumber k in flat space and thus similarly carries information about the angular scales $\theta \approx \pi/\ell$. Thus the maximum ℓ modes we can measure corresponds to the minimum angles we can measure with our experiment. The Planck CMB mission has angular resolution of $\theta_{\min} \approx 5' = 1.45 \times 10^{-3}$ radians and can thus measure modes upto $\ell_{\max} \approx \pi/\theta_{\min} \approx 2200$. The South Pole Telescope has 5 times better angular resolution ($\theta_{\min} \approx 1'$) and can thus measure modes up to $\ell_{\max} \approx 10^4$. In addition, as we saw in previous sections that the small scale modes are exponentially suppressed because of the Silk damping and thus in addition to high resolution, we also need high sensitivity in order to measure the small scales or high ℓ modes.

However, even before we get into experimental difficulties, we should solve a theoretical problem. In order to get theoretical predictions for modes upto some $\ell_{\max}$, we must solve the Boltzmann hierarchy, Eqs. 7.18, 7.54, 7.55, and 7.56, upto $\ell_{\max}$ complemented by few equations for other fluids and metric perturbations (Einstein's equations). Formally, each ℓ mode is coupled to the neighbouring $\ell - 1$ and $\ell + 1$ modes and we must solve the system of infinite number of coupled ODEs. We are saved by the fact that higher ℓ modes are suppressed due to Silk damping so we may truncate the hierarchy by approximating the modes beyond a cutoff mode ℓ_c by zero, $\Theta_\ell \approx 0$ for $\ell > \ell_c$. Any error we make at any ℓ will propagate to higher as well as smaller ℓ since all modes are coupled together and we are making a small error by assuming $\Theta_\ell = 0$ for $\ell > \ell_c$. If we want reasonable accuracy upto $\ell_{\max}$, we need to solve the hierarchy upto $\ell_c \sim \text{few} \times \ell_{\max}$. In principle there is no difficulty in solving a system of thousands of coupled ODEs on modern computers. However, our solutions are functions of cosmological parameters such as energy density of different components of the Universe Ω_i, Hubble constant H_0 etc. The minimal ΛCDM cosmological has six parameters and we must explore this 6 (or higher in extension of ΛCDM) dimensional parameter space using Monte-Carlo techniques in order to find the model/values of parameters that best fit the data and the error bars on the parameters or ideally the full probability density function (posterior) of the cosmological parameters. This implies that we must solve the Boltzmann equations millions of times varying the values of cosmological parameters. Thus we need a very fast way of solving for the *transfer functions* Θ_ℓ and solving thousands of coupled ODEs is not fast enough even on best processors available today.

An ingenious way was found by Seljak and Zaldarriaga [70] to solve the CMB Boltzmann equation which is fast enough to make Markov Chain Monte Carlo (MCMC) exploration of cosmological parameter space viable. All current CMB codes are based in this line of sight integration approach. The Boltzmann hierarchy we are trying to solve, Eqs. 7.18, 7.54, 7.55, and 7.56, were derived from the single Boltzmann equation 7.52. Instead of decomposing Eq. 7.52 into Legendre polynomials, we can try to solve the original Boltzmann equation directly. The idea is to first find the solution for $\Theta(k,\mu,\eta)$ and then decompose it into Legendre polynomials to get the solutions for $\Theta_\ell(k,\eta)$. We can rearrange Eq. 7.52 bringing all terms $\propto \Theta(k,\mu,\eta)$ to one side,

$$\begin{aligned}\Theta' + ik\mu\Theta - \tau_{\mathrm{T}}'\Theta &= -\phi' - ik\mu\psi - \tau_{\mathrm{T}}'\left[\Theta_0 + \mu v_{\mathrm{b}} - \frac{1}{2}P_2(\mu)\Pi\right]\\ &\equiv \tilde{S}(k,\mu,\eta),\end{aligned} \tag{7.105}$$

where we have also replaced Θ_2 with a new variable Π. Since we are ignoring polarization, $\Pi = \Theta_2$, but in general when we include polarization it will have contribution from the polarized part of the CMB also. We have also collected all terms on the right hand side into a source term $\tilde{S}$. If we had solutions for $\Theta_0, \Theta_2, \phi, \psi$, and v_{b}, then we can put them into the right hand side as source term and solve the above equation for $\Theta(k,\mu,\eta)$. Since we only need $\ell \le 2$ modes for the source term, it is sufficient to to solve the Boltzmann hierarchy, Eqs. 7.18, 7.54, 7.55, and 7.56 together with the equations for other fluid components and metric perturbations keeping terms upto $\ell_c \sim 10$ to give better than percent level accuracy. We can formally solve Eq. 7.105 by rewriting the left hand side as a total derivative giving,

$$e^{-ik\mu\eta+\tau_{\mathrm{T}}}\frac{\mathrm{d}}{\mathrm{d}\eta}\left(\Theta(k,\mu,\eta)e^{ik\mu\eta-\tau_{\mathrm{T}}}\right) = \tilde{S}(k,\mu,\eta) \tag{7.106}$$

and the solution is given by formally integrating from $\eta = 0$ to $\eta = \eta_0$,

$$\Theta(k,\mu,\eta)e^{ik\mu\eta-\tau_{\mathrm{T}}}\Big|_0^{\eta_0} = \int_0^{\eta_0}\mathrm{d}\eta e^{ik\mu\eta-\tau_{\mathrm{T}}}\tilde{S}(k,\mu,\eta). \tag{7.107}$$

From definition of τ_{T}, Eq. 7.15, as $\eta \to 0$, n_{e} and therefore $\tau_{\mathrm{T}} \to \infty$, while at $\eta = \eta_0$ we have $\tau_{\mathrm{T}} = 0$. Thus one of the boundary terms at $\eta = 0$ on the right hand side vanishes giving solution at η_0 as

$$\begin{aligned}\Theta(k,\mu,\eta_0)e^{ik\mu\eta_0} &= \int_0^{\eta_0}\mathrm{d}\eta e^{ik\mu\eta-\tau_{\mathrm{T}}}\tilde{S}(k,\mu,\eta)\\ \Theta(k,\mu,\eta_0) &= \int_0^{\eta_0}\mathrm{d}\eta e^{-ik\mu(\eta_0-\eta)}e^{-\tau_{\mathrm{T}}}\tilde{S}(k,\mu,\eta)\end{aligned} \tag{7.108}$$

Multiplying by Legendre polynomials and integrating over μ we get solution for $\Theta_\ell(k,\eta_0)$,

$$\begin{aligned}\Theta_\ell(k,\eta_0) &= \frac{1}{(-i)^\ell}\int_{-1}^{1}\frac{\mathrm{d}\mu}{2}P_\ell(\mu)\Theta(k,\mu,\eta_0)\\ &= \frac{1}{(-i)^\ell}\int_{-1}^{1}\frac{\mathrm{d}\mu}{2}P_\ell(\mu)\int_0^{\eta_0}\mathrm{d}\eta e^{-ik\mu(\eta_0-\eta)}e^{-\tau_\mathrm{T}}\tilde{S}(k,\mu,\eta). \qquad (7.109)\end{aligned}$$

We can do the integration over μ analytically. All terms with μ dependence in $\tilde{S}$ are $\propto P_1(\mu)=\mu$ and $P_2(\mu)=(3\mu^2-1)/2$ and we can get rid of the μ dependence in the source term by doing integration by parts. For any function $f(\eta)$ independent of μ, we have

$$\begin{aligned}\int_0^{\eta_0}\mathrm{d}\eta f(\eta)\mu e^{-ik\mu(\eta_0-\eta)} &= \int_0^{\eta_0}\mathrm{d}\eta\frac{f(\eta)}{ik\mu}\mu\frac{\mathrm{d}}{\mathrm{d}\eta}e^{-ik\mu(\eta_0-\eta)}\\ &= e^{-ik\mu(\eta_0-\eta)}\frac{f(\eta)}{ik}\Big|_0^{\eta_0}-\int_0^{\eta_0}\mathrm{d}\eta e^{-ik\mu(\eta_0-\eta)}\frac{\mathrm{d}}{\mathrm{d}\eta}\left(\frac{f(\eta)}{ik}\right)\\ &= -\int_0^{\eta_0}\mathrm{d}\eta e^{-ik\mu(\eta_0-\eta)}\frac{\mathrm{d}}{\mathrm{d}\eta}\left(\frac{f(\eta)}{ik}\right) \qquad (7.110)\end{aligned}$$

where in the second last line we have done integration by parts. One of the boundary terms vanished since our function $f\propto e^{-\tau_\mathrm{T}}$ and at $\eta=0$ $\tau_\mathrm{T}\to\infty$. We have ignored the second boundary term in the last line, since at $\eta=\eta_0$ we do not have any μ dependence left as the argument of the exponential vanished. This term will therefore contribute to only monopole today, $\Theta_0(\eta_0)$, which is not a frame independent observable and we are usually not interested in it. As we mentioned previously, we are usually interested in gauge invariant observables which are Θ_ℓ for $\ell\geq 2$ at linear order. Thus we are able to replace a factor of μ in the source term by a time derivative. Similarly for the P_2 term.

Exercise 30

Show by doing integration by parts twice that

$$\int_0^{\eta_0}\mathrm{d}\eta f(\eta)P_2(\mu)e^{-ik\mu(\eta_0-\eta)} = -\int_0^{\eta_0}\mathrm{d}\eta e^{-ik\mu(\eta_0-\eta)}\left(\frac{3}{2k^2}\frac{\mathrm{d}^2f(\eta)}{\mathrm{d}\eta^2}+\frac{1}{2}f(\eta)\right) \quad (7.111)$$

After using the above results, Eq. 7.110 and 7.111 for the source term, $\tilde{S}$, in Eq. 7.109 the only μ dependence that remains in Eq. 7.109 is in P_ℓ and the exponential.

The integral over μ is now easily done using the following identity

$$e^{-ik\mu(\eta_0-\eta)} = \sum_{\ell'}(2\ell'+1)i^{\ell'} j_{\ell'}\left[k(\eta-\eta_0)\right] P_{\ell'}(\mu), \tag{7.112}$$

where j_ℓ is the spherical Bessel function of first kind giving

$$\begin{aligned}\frac{1}{(-i)^\ell}\int_{-1}^{1}\frac{d\mu}{2}P_\ell(\mu)e^{-ik\mu(\eta_0-\eta)} &= \frac{1}{(-i)^\ell}\int_{-1}^{1}\frac{d\mu}{2}P_\ell(\mu)\sum_{\ell'}(2\ell'+1)i^{\ell'} j_{\ell'}\left[k(\eta-\eta_0)\right] P_{\ell'}(\mu)\\ &= (-1)^\ell j_\ell\left[k(\eta-\eta_0)\right] = j_\ell\left[k(\eta_0-\eta)\right]\end{aligned} \tag{7.113}$$

where we have used idenity $j_\ell(-x) = (-1)^\ell j_\ell(x)$ and orthogonality relation of Legendre polynomials Eq. 7.58. Putting everything together Eq. 7.109 becomes

$$\begin{aligned}\Theta_\ell(\eta_0,k) = \int_0^{\eta_0} d\eta\, j_\ell\left[k(\eta_0-\eta)\right]\Bigg[-\phi' e^{-\tau_T} + \frac{d}{d\eta}\left(\psi e^{-\tau_T}\right)\\ -\tau_T' e^{-\tau_T}\left(\Theta_0 + \frac{1}{4}\Pi\right) - \frac{d}{d\eta}\left(\frac{\tau_T' e^{-\tau_T} i v_b}{k}\right) - \frac{3}{4k^2}\frac{d^2}{d\eta^2}\left(\tau_T' e^{-\tau_T}\Pi\right)\Bigg]\end{aligned} \tag{7.114}$$

Defining the visibility function

$$g(\eta) = -\tau_T' e^{-\tau_T} \tag{7.115}$$

and rearranging terms we have

$$\begin{aligned}\Theta_\ell(\eta_0,k) &= \int_0^{\eta_0} d\eta\, j_\ell\left[k(\eta_0-\eta)\right]\Bigg[\left(\psi'-\phi'\right)e^{-\tau_T} + g(\eta)\left(\psi + \Theta_0 + \frac{1}{4}\Pi\right)\\ &\quad + \frac{d}{d\eta}\left(\frac{g(\eta) i v_b}{k}\right) + \frac{3}{4k^2}\frac{d^2}{d\eta^2}\left(g(\eta)\Pi\right)\Bigg]\\ &\equiv \int_0^{\eta_0} d\eta\, j_\ell\left[k(\eta_0-\eta)\right] S(k,\eta),\end{aligned} \tag{7.116}$$

where we have defined the source function $S(k,\eta)$. The visibility function $g(\eta)d\eta$ is the probability that a photon we are seeing last scattered in the time interval $d\eta$ at time η.

Exercise 31

Optional: Show that $g(\eta)$ has the correct normalization for a probability density function, i.e.

$$\int_0^{\eta_0} d\eta\, g(\eta) = 1 \tag{7.117}$$

The solution in Eq. 7.116 is a line of sight integral, as it integrates the source terms along the photon geodesics from the last scattering surface to us. The visibility function depends only on the recombination history of the Universe and can be calculated in advance. It is sharply peaked at the time of recombination. Before recombination, at $\eta < \eta_*$ the visibility function decays exponentially because of the $e^{-\tau_T}$ factor as $\tau_T \propto n_e$ increases as redshift increases or η decreases. As recombination progresses the electron number density and the optical depth drop and the visibility function $\propto -\tau_T' \propto n_e$ also drops rapidly. Once the first stars and galaxies form at $z \sim 10$, they flood the Universe with ionizing radiation reionizing the neutral atoms. The increase in electron density at reionization epoch results in a secondary smaller peak at $6 \lesssim z \lesssim 10$. Note that most of the source terms vanish when the visibility function vanishes, i.e. they constribute only when there is some (but not too much) scattering of photons by free electrons.

The spherical Bessel function does the geometric projection from the 3-D flat space to a 2-D spherical surface that we actually observe. The spherical Bessel function $j_\ell(x)$ has its peak when $\ell \approx x$. Each ℓ mode therefore gets contribution from a small range of k around $k \approx \ell/(\eta_0 - \eta)$. The transfer functions $\Theta_\ell(\eta_0, k)$ $\propto j_\ell$ are therefore similarly peaked around a narrow range of k. Together with the visibility function which is sharply peaked at $\eta = \eta_*$, the time of recombination, we see that the CMB anisotropies for mode ℓ that we see today would have contribution from $k \approx \ell/(\eta_0 - \eta_*) = \ell/r_{\rm LSS}$, where we have defined $r_{\rm LSS} = \eta_0 - \eta_*$ as the distance to the last scattering surface.

7.4.1 Integrated Sachs-Wolfe and Rees-Sciama effects

The first term in the square brackets in Eq. 7.116, $(\psi' - \phi')$ is not multiplied by the visibility function but just by $e^{-\tau_T}$ and contributes at all redshifts after recombination when τ_T becomes small $e^{-\tau_T}$ becomes approximately unity. This term involves time derivatives of the metric perturbations or gravitational potential and therefore contributes only when the gravitational potentials evolve. This term will therefore have contribution from superhorizon scales just after recombination when the Universe if transitioning from radiation domination to matter domination. As we saw earlier the gravitational potentials decay by $\approx 10\%$ during this transition and at the times of recombination the transition is not fully complete. This term is known as the integrated Sachs-Wolfe effect. This term again contributes at low redshifts when the Universe goes from matter domination to dark energy domination on large scales. In addition, on small scales, the gravitational potentials evolve because of the structure formation. Note that on large scales when we go from radiation to matter domination and again from matter to dark energy domination, the metric perturbations decay. On small scales on the other hand, non-linear gravitational collapse leads to amplification of the metric perturbations. The effect on small scales is therefore opposite of that on the large scales. On small scales, the effect of the same terms is known as the Rees-Sciama effect [71]. This effect can be understood as follow. When CMB photons are passing through a gravitational potential well, if the gravitational potential does not evolve during the time it takes the CMB photons to traverse it, the blue shift suffered by the photons as they enter the potential well is exactly cancelled by the redshift they suffer as they exit. If the graviational potentials change during this time however the bluehift and redshift do not cancel exactly leaving a net effect. For the ISW effect, since the potential well becomes weaker, redshift is smaller compared to blueshift as the photons traverse an overdensity and there is a net blueshift. The opposite happens when traversing a void. For the Rees-Sciama effect, also the redshift when traversing a collapsed overdensity is higher as the gravity has become stronger with the collapse and there is net redshift imprinted on the CMB photons.

7.4.2 Acoustic peaks and Doppler effect

Apart from the ISW terms, all other terms are multiplied by the visibility function. We can study the effect of these terms by approximating the visibility function by a Dirac delta function peaked at the recombination time η_*,

$$g(\eta) \approx \delta_{\mathrm{D}}(\eta - \eta_*). \tag{7.118}$$

Using this in Eq. 7.116, it is trivial to do the time integration giving (ignoring the subdominant Π terms)

$$\begin{aligned}\Theta_\ell(\eta_0, k) \approx{}& j_\ell\left[k(\eta_0 - \eta_*)\right]\left[\psi(k,\eta_*) + \Theta_0(k,\eta_*) + \frac{\mathrm{d}}{\mathrm{d}\eta}\frac{iv_{\mathrm{b}}(k,\eta)}{k}\bigg|_{\eta_*}\right] \\ &- \frac{\mathrm{d}}{\mathrm{d}\eta}\left(\frac{iv_{\mathrm{b}}(k,\eta)}{k} j_\ell\left[k(\eta_0-\eta)\right]\right)\bigg|_{\eta_*} \\ ={}& j_\ell\left[k(\eta_0-\eta_*)\right]\left[\psi(k,\eta_*) + \Theta_0(k,\eta_*)\right] + iv_{\mathrm{b}}(k,\eta_*) j'_\ell\left[k\left(\eta_0 - \eta_*\right)\right],\end{aligned} \tag{7.119}$$

where we have used the definition of derivative if Dirac delta function,

$$\int_0^{\eta_0} \delta'_D(\eta-\alpha) f(\eta)\mathrm{d}\eta = -f'(\alpha)\,, \alpha < \eta_0. \tag{7.120}$$

The Bessel function projects the 3-D perturbations into 2-D anisotropies on the sphere as discussed above. We see that the observed temperature anisotropies are proportional to the combination $\Theta_0(k,\eta_*) + \psi(k,\eta_*)$, as we claimed earlier in this chapter. The term $\propto v_{\rm b}$ is the Doppler effect and takes into acocunt the Doppler shift photons suffer as they are scattered into our direction by electrons moving towards or away from us. The factor of μ multiplying the $v_{\rm b}$ term has translated into the derivative of the spherical Bessel function and projects the radial velocity of the electrons, since this is the component of electron velocity that gives the Doppler shift to the photons coming to us.

We have already discussed the Sachs-Wolfe effect which comes from the $\Theta_0(k,\eta_*)+\psi(k,\eta_*)\approx(1/3)\psi(k,\eta_*)$ term on superhorizon scales at the time of recombination. This term is the picture of the Universe at the last scattering surface. On small scales we have acoustic oscillations and $\Theta_0(k,\eta_*)+\psi(k,\eta_*)$ has extrema at $kr_{\rm s}\approx n\pi$ for $n\geq 1$ an integer or $k\approx n\pi/r_{\rm s}$. The Bessel function projects these peaks to ℓ modes given by

$$\ell\approx k(\eta_0-\eta_*)=\frac{r_{\rm LSS}}{r_{\rm s}}n\pi \tag{7.121}$$

Exercise 32

Using Eq. 7.121 find the approximate position (ℓ) of the first acoustic peak of the CMB. Use Planck mission cosmological parameters and redshift of recombination, $z_*=1100$. What is the angular scale of the first peak on our sky ?

Note that the correspondence between the k and ℓ is not one to one. A range of k modes will contribute to a particular ℓ mode when we take the Fourier transform below of $\Theta_\ell(k,\eta_0)$ to get the anisotropy observed by us in real space. This has the effect of smoothing out the oscillations a little making the peaks broader and smaller and raising the troughs in the power spectrum. Also note that even though in acoustic oscillations the velocity is out of phase with temperature perturbations, the Doppler term enters as gradient of velocity which is again in phase with Θ_0. This is because we have to project the radial component of velocity and velocity is in same direction as $\mathbf{k}$. The Doppler effect results in a small increase in the amplitude of observed CMB anisotropies and it is not fully in-phase with Θ_0, even with the time derivative resulting in filling up of the troughs and reducing the contrast between the troughs and the peaks.

7.4.3 Reionization

Once the Universe has become neutral, the radiation pressure disappears and baryons can finally collapse to form the first stars and galaxies. In ΛCDM cosmology, we expect the first stars and galaxies to start forming at $z\sim 30$. These first stars and galaxies emit UV and X-ray photons which start the process of ionizing the Universe. We expect the reionization to start becoming significant at $z\sim 10$ and the Universe to become ionized almost completely (hydrogen and first ionization of helium) by $z \sim 6$ with only small pockets of neutral gas remaining in dense collapse structure i.e. galaxies.

The free electrons created during the reionization will scatter the CMB photons free streaming to us from the last scattering surface. We can define the reionization optical depth by integrating the Thomson optical depth from the last scattering surface until today,

$$\tau_{\rm ri} = -\int_{\eta_{\rm ri}}^{\eta_0} {\rm d}\eta \tau'_{\rm T} = \int_{\eta_{\rm ri}}^{\eta_0} {\rm d}\eta \, n_{\rm e}\sigma_{\rm T} a, \tag{7.122}$$

where $\eta_{\rm ri}$ is the conformal time value of conformal time at a point between the end of recombination and before first stars form and begin reionization, an epoch known as the *dark ages*. The precise value of $\eta_{\rm ri}$ is not important as long as it is taken to be some time during the dark ages.

We can define the Thomson optical depth at a time $\eta < \eta_{\rm ri}$ as the sum of $\tau_{\rm ri}$ and the optical depth from $\tau_{\rm ri}$ to η,

$$\begin{aligned} \tau_{\rm T}(\eta) &= -\int_{\eta}^{\eta_0} {\rm d}\eta \tau'_{\rm T} \\ &= -\int_{\eta}^{\eta_{\rm ri}} {\rm d}\eta \tau'_{\rm T} - \int_{\eta_{\rm ri}}^{\eta_0} {\rm d}\eta \tau'_{\rm T} \\ &\equiv \tilde{\tau_{\rm T}} + \tau_{\rm ri}, \end{aligned} \tag{7.123}$$

where in the last line we have defined $\tilde{\tau_{\rm T}}$. In the absence of reionization, $\tilde{\tau_{\rm T}} = \tau_{\rm T}$. We also define the visibility function in the absence of reionization,

$$\tilde{g}(\eta) = -\tau'_{\rm T} e^{-\tilde{\tau_{\rm T}}}, \tag{7.124}$$

so that the full visibility function is

$$g(\eta) = \tilde{g}(\eta) e^{-\tau_{\rm ri}} \tag{7.125}$$

Similarly, we can break up the integral in the line of sight solution, Eq. 7.116 into two integrals, one upto reionization time, $\eta_{\rm ri}$ and the other from reionization until today. To do this, we note that the all terms in the integrand are instantaneous terms which depend only upon the instantaneous values of variables and are therefore not affected in anyway on breaking up the integral. The only exception is the optical depth $\tau_{\rm T}(\eta)$ which itself is an integral from time η until η_0 and will therefore be different if there is no reionization. We also note that the Thomson optical depth only enters in the exponential factor, $e^{-\tau_T} = e^{-\tilde{\tau}_T}e^{-\tau_{ri}}$, which is common to all terms in $S(k,\eta)$. We can therefore write the source term as

$$S(k,\eta) = S^{\text{no-ri}}(k,\eta)e^{-\tau_{ri}} \tag{7.126}$$

where the source term for the no-reionization case is,

$$\begin{aligned} S^{\text{no-ri}}(k,\eta) = (\psi' - \phi')\, e^{-\tilde{\tau}_T} + \tilde{g}(\eta)\left(\psi + \Theta_0 + \frac{1}{4}\Pi\right) \\ + \frac{d}{d\eta}\left(\frac{\tilde{g}(\eta) i v_b}{k}\right) + \frac{3}{4k^2}\frac{d^2}{d\eta^2}\left(\tilde{g}(\eta)\Pi\right). \end{aligned} \tag{7.127}$$

The line of sight solution can therefore be written as

$$\begin{aligned} \Theta_\ell(\eta_0,k) &= \int_0^{\eta_{ri}} d\eta\, j_\ell\left[k(\eta_0-\eta)\right] S(k,\eta) + \int_{\eta_{ri}}^{\eta_0} d\eta\, j_\ell\left[k(\eta_0-\eta)\right] S(k,\eta) \\ &= \int_0^{\eta_{ri}} d\eta\, j_\ell\left[k(\eta_0-\eta)\right] e^{-\tau_{ri}} S^{\text{no-ri}}(k,\eta) + \int_{\eta_{ri}}^{\eta_0} d\eta\, j_\ell\left[k(\eta_0-\eta)\right] S(k,\eta) \\ &= e^{-\tau_{ri}}\Theta_\ell^{\text{no-ri}}(\eta_0,k) + \int_{\eta_{ri}}^{\eta_0} d\eta\, j_\ell\left[k(\eta_0-\eta)\right] S(k,\eta) \end{aligned} \tag{7.128}$$

The Eq. 7.128 is the same solution as Eq. 7.116, with the contribution to the CMB anisotropies from reionization separated from the CMB anisotropies originating at the last scattering surface or before reionization.

The modification of CMB anisotropies due to reionization, that we are discussing in this section, allows us to constrain τ_{ri} from observations of CMB anisotropies. The measurements by Planck CMB space mission [72] imply $\tau_{ri} \approx 0.054$. The first term in Eq. 7.128 is just the CMB anisotropies we would see in the absence of reionization multiplied by a damping factor of $e^{-\tau_{ri}}$. Thus reionization decreases the amplitude of the CMB anisotropies originating at the last scattering surface by $\approx$ 5.5%, blurring the picture of the last scattering surface that we see. The second term includes the new anisotropies generated due to Thomson scattering during reionization as well the ISW effect later during transition to dark energy dominated era from matter dominated era.

During reionization, we are deep in the matter dominated era and ISW effect is negligible. Since the mean free path of the CMB photons is of horizon size, the CMB temperature perturbations, Θ_0, are erased below horizon scale at reionization, η_{ri}. The Doppler term in $S(k,\eta)$ ($\propto v_b$) is thus the most dominant term. However, since reionization is an extended epoch, the contribution to the Doppler term from different regions along the line of sight would give blue shift or red shift randomly. For small scales, there will be numerous regions with randomly oriented baryon velocity $v_{\rm b}$, resulting in cancellation of blue shifts and redshifts. The number of independent regions at scale $\lambda = 2\pi$ in a box of length L is $\sim L/\lambda$. For reionization L corresponds to the distance along the line of sight from approximately beginning of reionization to the end of reionization. After the end of reionization, the optical depth decreases at $n_{\rm e} \propto a^{-3}$ and contribution to CMB anisotropies from Thomson scattering is very small. Thus For large scales, there will be smaller number of independent regions along the line of sight and cancellation less severe. Similar arguments apply to the ψ term.

Thus on small scales the first damping term in Eq. 7.128 would be dominant resulting a suppression of anisotropies. On large scales the primary CMB anisotropies from the last scattering surface consist of mostly the Sachs-Wolfe effect which is much smaller compared to the acoustic peaks on scales of sound horizon at recombination. Thus, suppression of primary anisotropies, which are already small on large scales, is subdominant compared to the creation of new anisotropies from Thomson scattering and the second term dominates. As during recombination, the amplitude of new anisotropies is maximum on horizon scale at reionization, $k \approx 1/\eta_{\rm ri}$. On scale larger than the horizon size we just have the Sachs-Wolfe effect while on scales of order horizon size we have the contribution from velocities $v_{\rm b}$ matter falling into gravitational potentials. This infall also creates quadrupolar anisotropies, in the rest frame of free electrons, due to coupling of $\ell = 1$ modes with $\ell = 2$ modes in Boltzmann hierarchy, similar to the generation of quadrupole during recombination. This quadrupole on horizon scale creates new polarization in CMB on scales $k \approx 1/\eta_{\rm ri}$. Thus we expect a peak or bump in both CMB temperature anisotropies and polarization anisotropies on scales corresponding to the horizon size at reionization. Thus reionization bump is the dominant feature through which we measure the optical depth to reionization, $\tau_{\rm ri}$.

Exercise 33

Assume that reionizatio happens at redshift $z_{\rm ri} \approx 8$. At what ℓ will we see the reionization bump today ? You can use the cosmological parametrs from Planck 2018 paper [72].

7.5 From initial conditions and transfer functions to CMB anisotropies

The solution that we have derived so far are called *transfer functions*. In linear evolution the final solution is given by multiplying the transfer functions by the initial conditions which we had labeled by A. The reansfer functions, we had argued as well as directly seen from the evolution equations, do not depend on the absolute direction of Fourier mode, $\hat{\mathbf{k}}$, but only the relative direction, $\mu = \hat{\mathbf{k}}.\hat{\mathbf{n}}$ and magnitude of Fourier mode, k, for photons (and other relativistic free streaming species such as neutrinos) and only on the magnitude k for fluids such as baryons and dark matter. Nothing however stops us from having different initial amplitudes for different modes. In fact, initial amplitudes will depend on the physical process which gave rise to the perturbations in the early Universe and would in general be a function of wavevector $\mathbf{k}$, i.e. $A = A(\mathbf{k})$.

In the inflation theory of generation of initial conditions, the initial fluctuations are a realization of a isotropic Gaussian random field. This means that the statistical properties such as the correlation functions, which are ensemble averages of a large number of realizations of the field, will not be a function of direction $\hat{\mathbf{k}}$. However, any particular realization will have randomly chosen initial amplitudes $A(\mathbf{k})$ which will be different for different directions $\hat{\mathbf{k}}$ for same k. It is only in a statistical average sense that they do not depend on the direction. In any case, to completely specify the initial density field, $A(\mathbf{x})$, we need information about all Fourier modes $A(\mathbf{k})$.

The perturbations in any quantity, including the CMB anisotropes, at any time is given by the product of transfer function with initial amplitude. Thus the full solution for CMB anisotropies is given by

$$\Theta_\ell(\mathbf{k}, \eta) = \Theta_\ell(k, \eta) A(\mathbf{k}), \tag{7.129}$$

where we are using the same symbol for full solution as the transfer functions but they are distinguishable by their arguments. Similarly, we define dark matter density field at any time η also by

$$\delta(\eta, \mathbf{k}) = \delta(\eta, k) A(\mathbf{k}), \tag{7.130}$$

where we absorb any numerical factors into the definition of transfer function $\delta(\eta, k)$, so that at $\eta = 0$ the adiabatic conditions, Eq. 7.32, are satisfied.

We want to connect the solution today, $\Theta_\ell(\mathbf{k}, \eta_0)$, with the temperature anisotropy in real space that we observe today, $\Theta(\mathbf{x}_0, \eta_0, \hat{\mathbf{n}})$ from our comoving spacetime position at $\mathbf{x}_0, \eta_0$ in direction $\hat{\mathbf{n}}$. This is easily done by first doing the inverse transformation in angular space and using addition theorem of spherical harmonics Eq. 7.11, giving,

$$\begin{aligned}\Theta(\mathbf{k}, \eta_0, \hat{\mathbf{n}}) &= \sum_\ell (-i)^\ell (2\ell + 1) \Theta_\ell(\mathbf{k}, \eta_0) P_\ell(\hat{\mathbf{k}}.\hat{\mathbf{n}}) \\ &= 4\pi \sum_{\ell,m} (-i)^\ell \Theta_\ell(k, \eta_0) A(\mathbf{k}) Y^*_{\ell m}(\hat{\mathbf{k}}) Y_{\ell m}(\hat{\mathbf{n}})\end{aligned} \tag{7.131}$$

Taking the Fourier transform gives us the real space temperature anisotropies which we evaluate at our position $\mathbf{x}_0$,

$$\begin{aligned}\Theta(\mathbf{x}_0,\eta_0,\hat{\mathbf{n}}) &= \int \frac{\mathrm{d}k^3}{(2\pi)^3}e^{i\mathbf{k}.\mathbf{x}_0}\Theta(\mathbf{k},\eta_0,\hat{\mathbf{n}})\\ &= \int \frac{\mathrm{d}k^3}{(2\pi)^3}e^{i\mathbf{k}.\mathbf{x}_0}4\pi\sum_{\ell,m}(-i)^{\ell}\Theta_{\ell}(k,\eta_0)A(\mathbf{k})Y^*_{\ell m}(\hat{\mathbf{k}})Y_{\ell m}(\hat{\mathbf{n}}) \qquad (7.132)\end{aligned}$$

and taking the spherical harmonic transform gives the expression for spherical harmonic coefficients at our position,

$$\begin{aligned}a_{\ell m}(\mathbf{x}_0,\eta_0) &= \int \mathrm{d}^2\hat{\mathbf{n}}\ Y^*_{\ell m}(\hat{\mathbf{n}})\Theta(\mathbf{x}_0,\eta_0,\hat{\mathbf{n}})\\ &= \int \mathrm{d}^2\hat{\mathbf{n}}\ Y^*_{\ell m}(\hat{\mathbf{n}})\int \frac{\mathrm{d}^3k}{(2\pi)^3}e^{i\mathbf{k}.\mathbf{x}_0}4\pi\sum_{\ell',m'}(-i)^{\ell'}\Theta_{\ell'}(k,\eta_0)A(\mathbf{k})Y^*_{\ell' m'}(\hat{\mathbf{k}})Y_{\ell' m'}(\hat{\mathbf{n}}),\end{aligned}$$
(7.133)

where $\mathrm{d}^2\hat{\mathbf{n}} \equiv \mathrm{d}\Omega \equiv \mathrm{d}\sin\theta\mathrm{d}\theta\mathrm{d}\phi$ represents the two dimensional differential solid angle with θ the spherical polar coordinate and ϕ the azimuthal spherical coordinate in the spherical polar coordinate system centered at the observer at $\mathbf{x}_0,\eta_0$. This integral over angles is trivially done using orthogonality of spherical harmonics, Eq. 7.9 giving,

$$\begin{aligned}a_{\ell m}(\mathbf{x}_0,\eta_0) &= \int \frac{\mathrm{d}^3k}{(2\pi)^3}e^{i\mathbf{k}.\mathbf{x}_0}4\pi\sum_{\ell',m'}(-i)^{\ell'}\Theta_{\ell'}(k,\eta_0)A(\mathbf{k})Y^*_{\ell' m'}(\hat{\mathbf{k}})\delta_{\ell\ell'}\delta_{mm'}\\ &= 4\pi\int \frac{\mathrm{d}^3k}{(2\pi)^3}e^{i\mathbf{k}.\mathbf{x}_0}(-i)^{\ell}Y^*_{\ell m}(\hat{\mathbf{k}})\Theta_{\ell}(k,\eta_0)A(\mathbf{k}) \qquad (7.134)\end{aligned}$$

7.5.1 Gaussian random fields

The observed CMB anisotropies and hence the spherical harmonic coefficients, $a_{\ell m}(\mathbf{x}_0,\eta_0)$ depend on the position of the observer $\mathbf{x}_0$. This is not a surprise. The CMB sky observed by a far away observer, e.g. in a galaxy in the Coma cluster, would be different from what is observed by us. It is a random realization of a random field due to the initial field $A(\mathbf{k})$. Thus the exact CMB sky observed by us is not predictable. What is predictable, from a theory such as inflation, is the power spectrum of the initial random field, $A(\mathbf{k})$ and hence the power spectrum of the CMB temperature anisotropies or two point correlation function, $\langle a_{\ell_1 m_1}a^*_{\ell_2 m_2}\rangle$, where the angular brackets indicate ensemble average over infinitely many realizations of the random field. In particular, the initial perturbations $A(\mathbf{k})$ or $A(\mathbf{x})$ predicted by inflation are Gaussian and are statistically isotropic and homogeneous.

Gaussianity means that the probability distribution of the field at any point $\mathbf{x}_1$ is given by the Gaussian distribution, $P(A(\mathbf{x}_1)) = \mathcal{G}(0, \sigma(\mathbf{x}_1))$ with zero mean and same variance $\sigma(\mathbf{x}_1)$. This is not all. The joint probability distribution for the value of field at two points, $P(A(\mathbf{x}_1)A(\mathbf{x}_2))$ is a two dimensional Gaussian distribution with covariance $C(\mathbf{x}_1, \mathbf{x}_2)$. Similary the joint probability distribution for the value of field at n points is a n-dimensional Gaussian distribution. The homogeniety and isotropy of the field implies that the covariances, σ, C etc. do not depend on the absolute position or orientation of the points but only the relative distance between the points. In terms of the correlation functions this means that the two point correlation function depends only on the distance, $r = |\mathbf{x}_1 - \mathbf{x}_2|$ between the two points,

$$\langle A(\mathbf{x}_1)A(\mathbf{x}_2)\rangle = C(|\mathbf{x}_1 - \mathbf{x}_2|) = C(r). \tag{7.135}$$

Note that homogeneity implies that the correlation function depends only on relative position, $\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2$ while the isotropy implies that the correlation function in addition does not depend on the orientation of $\mathbf{r}$ but only on its magnitude r. Note that the ensemble average indicated by angular brackets $\langle\cdot\rangle$ acts only on the random components of the integrand, i.e. the random fields A. In Fourier space, the Fourier transform of the correlation function is the power spectrum. In Fourier space we have

$$\begin{aligned}\langle A(\mathbf{k})A^*(\mathbf{k}')\rangle &= \left\langle \int \mathrm{d}^3x A(\mathbf{x})e^{-i\mathbf{k}.\mathbf{x}} \int \mathrm{d}^3x' A(\mathbf{x}')e^{i\mathbf{k}'.\mathbf{x}'}\right\rangle \\ &= \int \mathrm{d}^3x \int \mathrm{d}^3r \,\langle A(\mathbf{x})A(\mathbf{x}+\mathbf{r})\rangle\, e^{-i\mathbf{k}.\mathbf{x}}e^{-i\mathbf{k}'.(\mathbf{x}+\mathbf{r})},\end{aligned} \tag{7.136}$$

where in the second line we have done a change of variables from $\mathbf{x}'$ to $\mathbf{r}$ defined by $\mathbf{x}' = \mathbf{x} + \mathbf{r}$ which is just a translation of coordinates. Note that $\mathbf{x}$ is kept constant in the inner integral when doing a change of variables. Using Eq. 7.135 we get

$$\begin{aligned}\langle A(\mathbf{k})A^*(\mathbf{k}')\rangle &= \int \mathrm{d}^3x \int \mathrm{d}^3r C(r)e^{-i\mathbf{k}.\mathbf{x}}e^{i\mathbf{k}'.(\mathbf{x}+\mathbf{r})} \\ &= \int \mathrm{d}^3r C(r)e^{-i\mathbf{k}'.\mathbf{r}} \int \mathrm{d}^3x e^{i(\mathbf{k}'-\mathbf{k}).\mathbf{x}} \\ &= \int \mathrm{d}^3r C(r)e^{-i\mathbf{k}'.\mathbf{r}}(2\pi)^3\delta_D(\mathbf{k}'-\mathbf{k}) \\ &\equiv (2\pi)^3\delta_D(\mathbf{k}'-\mathbf{k})P_A(k),\end{aligned} \tag{7.137}$$

where we have used the definition of the Dirac delta function to do the integration over $\mathbf{x}$ and defined the power spectrum as Fourier transform of real space 2-point correlation function. The power spectrum $P_A(k)$ only depends on the magnitude of k (isotropy) and vanishes if $\mathbf{k}' \neq \mathbf{k}$ (homogeneity). Thus homogeneity of the random field implies that different Fourier modes are independent. Also, the Guassianity of the random field means that the 2-point correlation function or the power spectrum contain all statistical information. In particular, higher order correlation functions can be written in terms of two point correlation functions using Wick's theorem and the n-point correlation functions vanish for n a odd number.

We can now calculate the angular power spectrum, $\langle a_{\ell m} a^*_{\ell' m'} \rangle$. From Eq. 7.134 we get

$$\begin{aligned}
\langle a_{\ell m} a^*_{\ell' m'} \rangle &= \left\langle (4\pi)^2 \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \int \frac{\mathrm{d}^3 k'}{(2\pi)^3} \right. \\
&\left. e^{i\mathbf{k}.\mathbf{x}_0} (-i)^\ell Y^*_{\ell m}(\hat{\mathbf{k}}) \Theta_\ell(k,\eta_0) A(\mathbf{k}) e^{-i\mathbf{k}'.\mathbf{x}_0} (i)^{\ell'} Y_{\ell' m'}(\hat{\mathbf{k}}') \Theta_{\ell'}(k',\eta_0) A^*(\mathbf{k}') \right\rangle \\
&= (4\pi)^2 \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \int \frac{\mathrm{d}^3 k'}{(2\pi)^3} e^{i(\mathbf{k}-\mathbf{k}').\mathbf{x}_0} (-i)^\ell i^{\ell'} Y^*_{\ell m}(\hat{\mathbf{k}}) Y_{\ell' m'}(\hat{\mathbf{k}}') \Theta_\ell(k,\eta_0) \Theta_{\ell'}(k',\eta_0) \\
&(2\pi)^3 P_A(k) \delta_D(\mathbf{k}' - \mathbf{k}) \\
&= (4\pi)^2 \int \frac{\mathrm{d}^3 k}{(2\pi)^3} (-i)^\ell i^{\ell'} Y^*_{\ell m}(\hat{\mathbf{k}}) Y_{\ell' m'}(\hat{\mathbf{k}}) \Theta_\ell(k,\eta_0) \Theta_{\ell'}(k,\eta_0) P_A(k) \\
&= \frac{2}{\pi} \int \mathrm{d}k\, k^2 \Theta^2_\ell(k,\eta_0) P_A(k) \delta_{\ell\ell'} \delta_{mm'} \\
&\equiv C_\ell(\eta_0) \delta_{\ell\ell'} \delta_{mm'},
\end{aligned} \tag{7.138}$$

where we used the Dirac delta function to do the k' inetgral and in the last steps we see that only angular dependence that is left is in $Y_{\ell m}$. We used orthogonality of spherical harmonics to do the angular integral and are left with only one k integration. Note that the dependence on $\mathbf{x}_0$ has vanished and thus all observers will measure the same power spectrum C_ℓ defined in the last step,

$$C_\ell(\eta_0) = \frac{2}{\pi} \int \mathrm{d}k\, k^2 \Theta^2_\ell(k,\eta_0) P_A(k) \tag{7.139}$$

as is expected for a homogeneous and isotropic Universe. Also note that the homogeneity implies that the correlations between the $a_{\ell m}$ vanish unless $\ell = \ell'$ and $m = m'$ and isotropy implies that C_ℓ does not depend on m. We see from Eq. 7.139 that C_ℓ is an integral over all k modes and each ℓ modes therefore gets contribution from all k modes. The geometric factor of spherical Bessel function j_ℓ in Θ_ℓ does the projection from k space to ℓ space. Since $j_\ell[k(\eta_0 - \eta)]$ is peaked at $\ell \approx k(\eta_0 - \eta)$, the anisotropies generated at any time η contribute to each ℓ mode dominantly from a small range of modes around $k \approx \ell/(\eta_0 - \eta)$. The visibility function ensures that the dominant contribution comes from around $\eta \approx \eta_*$, the last scattering surface. The transfer functions are usually normalized so that A represents the metric perturbation ϕ in Newtonian gauge or the curvature perturbation $\mathcal{R}$ defined in the comoving gauge, in which case the initial power spectrum $P_A(k)$ is the power spectrum of the initial curvature perturbation.

We could have done our analysis in real space also, calculating the correlation function of the temperature field, $\Theta(\mathbf{x}_0, \eta_0, \hat{\mathbf{n}})$. The main advantage of doing analysis in spherical harmonic space is that the two-point function or the power spectrum is diagnoal in harmonic space, as is evident from Eq. 7.138 or Eq. 7.137. Thus different Fourier or ℓ modes are uncorrelated. In the real space, Eq. 7.135, shows that the covariance between any two points will be non-zero and depend on the distance between them. In other words, the spherical harmonic coefficients, $a_{\ell m}$ are Gaussian distributed, for a Gaussian random field, and the covariance of joint probability distribution of all $a_{\ell m}$ is given by Eq. 7.138, which is diagonal. In real space, the diagonal entries give the variance of fluctuations at each point, and because of statistical homogeneity, all diagnoal entries are same. The off diagnoal elements give covariance between fluctuations at different spatial positions, which are also non-zero and equal to $C(r)$. For statistical analysis, it is convenient and computationally efficient to deal with diagonal matrices and hence with $a_{\ell m}$.

7.5.2 Observations: ergodicity and cosmic variance

We have defined the power spectrum above as an ensemble average over infinite realizations of universes. This is easy to do theoretically. Observationally we have however access to a single realization of the Universe, the one in which we exist. We however cannot predict theoretically what a single realization of Universe would look like. We can however still calculate statistical quantities, such as the angular power spectrum C_ℓ, from observations of a single realization of the Universe under certain assumptions. Under certain hypothesis we can extract many estimates of the statistical quantity of interest from observational data and then take average of these estimates. In frequentist statistics terminology such a quantity is called an *estimator* of the actual quantity we are interested in.

One assumption we can make when working with a large volume of the Universe is the *ergodic hypothesis*. The ergodic hypothesis says that for quantities which are local or require a small volume of the Universe to get a single measurement, we can treat *sufficiently* distant different parts of the Universe as independent realizations and measurements of the statistical quantity of interest in different parts of the Universe or *different parts of data* can be treated as independent measurements. We can therefore replace the ensemble average with an average of these independent measurements.

Ergodic theorem: The ergodic theorem states that the average of measurements in well separated parts of the Universe (or data) approaches the ensemble average if the Universe is statistically homogeneous or stationary. A formal proof of ergodic theorem can be found in Appendix D of [25].

For C_ℓ we have already proven that $\langle a_{\ell m} a^*_{\ell m}\rangle$ is independent of m for an isotropic Universe (i.e. stationarity on a sphere) and that $\langle a_{\ell m} a^*_{\ell m'}\rangle$ vanishes for $m \neq m'$. We can therefore take the product $a_{\ell m} a^*_{\ell m}$ for each m as an independent measurement/realization of C_ℓ. For each ℓ we therefore have $2\ell + 1$ independent measurements of C_ℓ. We define our estimator of C_ℓ as the avearage of these independent measurements,

$$\hat{C}_\ell = \frac{1}{2\ell + 1} \sum_m a_{\ell m} a^*_{\ell m}. \tag{7.140}$$

At high values of ℓ, we will have a large number of independent measurements and thus an accurate estimate of C_ℓ while for small values of ℓ we have only a few independent measurements and so the errorbar on our estimate of C_ℓ will be large.

We can quantify and calculate the *average* squared error we should expect on C_ℓ by taking the ensemble average of squared difference or variance (σ^2_ℓ) between our estimate of $\hat{C}_\ell$ and true C_ℓ,

$$\sigma^2_\ell = \left\langle \left(\hat{C}_\ell - C_\ell\right)^2 \right\rangle \tag{7.141}$$

The ensemble average of our estimator is actually the true power spectrum, $\langle \hat{C}_\ell \rangle = C_\ell$, therefore

$$\begin{aligned}
\sigma^2_\ell &= \left\langle \hat{C}_\ell \hat{C}_\ell \right\rangle - C^2_\ell \\
&= \frac{1}{(2\ell + 1)^2} \sum_{mm'} \left\langle a_{\ell m} a^*_{\ell m} a_{\ell m'} a^*_{\ell m'} \right\rangle - C^2_\ell \\
&= \frac{1}{(2\ell + 1)^2} \sum_{mm'} \left[\left\langle a_{\ell m} a^*_{\ell m} \right\rangle \left\langle a_{\ell m'} a^*_{\ell m'} \right\rangle + \left\langle a_{\ell m} a^*_{\ell m'} \right\rangle \left\langle a_{\ell m} a^*_{\ell m'} \right\rangle \right. \\
&\left. + \left\langle a^*_{\ell m} a^*_{\ell m'} \right\rangle \langle a_{\ell m} a_{\ell m'} \rangle \right] - C^2_\ell \\
&= \frac{1}{(2\ell + 1)^2} \left[C^2_\ell (2\ell + 1)^2 + C^2_\ell (2\ell + 1) + C^2_\ell (2\ell + 1) \right] - C^2_\ell \\
&= \frac{2}{2\ell + 1} C^2_\ell
\end{aligned} \tag{7.142}$$

where we have used Isserlis theorem, also known as Wick's theorem in particle physics, to write the 4-point ensemble average of Gaussian random variables, $a_{\ell m}$ as sum of products of 2-point ensemble averages. Thus the minimum statistical error on C_ℓ, assuming ideal measurements, is

$$\sigma_\ell = \sqrt{\frac{2}{2\ell+1}} C_\ell. \tag{7.143}$$

The error on $\hat{C}_\ell$ is therefore proportional to C_ℓ. This irreducible error on cosmological measurements, coming from the fact that we have access to only one realization of the Universe, is called *cosmic variance*. We are limited in precision by cosmic variance in measurements of statistical properties of all cosmic fields, such as the dark matter power spectrum or higher order order correlation functions.

Note that since we do not have access to the true C_ℓ, but only the estimate $\hat{C}_\ell$, we do not know the *true* error σ_ℓ but can only calculate an *estimate* $\hat{\sigma}_\ell$ of the error. The error is therefore also known with finire accuracy and we can calculate an estimate of the error on our error estimate i.e. error on σ_ℓ and so on.

If we do not have access to the full sky, which is the case in reality, we cannot independently measure all m modes for a given ℓ and the error on C_ℓ is larger by an amount equal to $1/\sqrt{f_{\rm sky}}$, where $f_{\rm sky}$ is the sky fraction used. For example, in the directions towards the plane of our Galaxy, we have significant emissions from dust and molecular clouds in the CMB frequency bands and we cannot reliably distinguish between the two. The region around the Galactic plane is therefore not used for cosmological analysis. A nice analytical analysis procedure for the case of fractional sky is give in [73].

Chapter 8

Initial conditions and Inflation

Cosmological observations over the past 50 years have built up a picture of the Universe which is extremely simple. The Universe can be understood in terms of a few numbers describing its large scale average structure and follows terrestrial laws of physics upto very early times. In fact, in some sense, the simplest possible Universe that we can imagine, which is consistent with known laws of physics, seems to be the one we are living in. There are a few complications such as matter-antimatter asymmetry (or baryogenesis), dark matter, and dark energy, but these problems are not much more complez than other problems with the standard model of particle physics.

The question then arises: why is the Universe so simple ? In fact the further we go in the past, simpler the Universe becomes in a sense we will define precisely below.

We can extrapolate in the past only as far as the Planck scale defined by

$$E_{\mathrm{P}} = \left(\frac{\hbar c^5}{G}\right)^{1/2} = 10^{19}\ \mathrm{GeV}. \tag{8.1}$$

This is the energy scale, i.e. energy density of $\sim {E_{\mathrm{P}}}^4$, where both quantum effects and gravity important and classical gravity can no longer be used. Extrapolation with classical gravity takes us past the Planck scale to a singularity. Thus with classical gravity, the earliest we can start our calculations ia at Planck scale, $\sim 10^{19}$ GeV. Today the energy scale is given by the temperature of the CMB $\approx 10^{-4}$ eV (2.725 K). Therefore we must specify initial conditions at a scale factor of a_{P} given by

$$\frac{a_{\mathrm{P}}}{a_0} = \frac{10^{-4}\ \mathrm{eV}}{10^{19}\ \mathrm{GeV}} = 10^{-32} \tag{8.2}$$

after which we can evolve the Universe using classical gravity. To be more precise, when the energy density is $\left(10^{19}\right)^4$ GeV^4, we have energy of the order 10^{19} GeV confined to a volume of size

$$L_{\mathrm{P}}{}^3 = \left(\sqrt{\frac{\hbar G}{c^3}}\right)^3 \approx \left(10^{-33}\ \mathrm{cm}\right)^3, \tag{8.3}$$

where L_{P} is the Planck length. Quantum effects are important because of small scales and gravity is important because of large curvature.

8.1 A simple Universe and problems with it

The simple Universe, demanded by cosmological observations interpreted with classical Einstein gravity and standard model of particle physics, however poses a number of problems related to the initial conditions. The efforts to find a solution to these problems lead us to the theory of inflation. We will see that inflation solves some problems, makes others less severe, and raises new questions about itself. A good reference for inflation is the textbook by Mukhanov [74].

8.1.1 Horizon problem

The first indication of a problem comes from the CMB. The comoving horizon size at recombination is $\eta_* = 280$ Mpc. The distance of the last scattering surface from us is $\eta_0 - \eta_* \approx \eta_0 = 14 \times 10^3$ Mpc. The horizon problem is sketched in Fig. 8.1. At the time of recombination, the patches of sky separated by angles $\theta \gg 280/14000 \approx 1^o$ as seen by us today were out of causal contact if we simply extrapolate the current Universe to $t = 0$. However, we see that the CMB temperature in these widely separated patches of the sky has the same value of 2.725 K, with very small differences in CMB temperature in different directions of order $\approx 10^{-4}$ K. The inhomogeneities in the matter distribution in different directions are also small. The Universe thus has almost exactly the same temperature and composition in the causally disconnected parts of the Universe at the time of recombination. We have also seen that initial fluctuations in the the energy density of different constituents either grows or oscillates or remains constant. The initial fluctuations, with the exception of Silk damping which can be taken into account, do not decay with expansion. Thus, the perturbations in the Universe at the time of recombination are of same order as at $t \to 0$. This means that if we want to set our initial conditions at a_{P} or $t_{\mathrm{P}} = \sqrt{\hbar G/c^5} = 5 \times 10^{-44}$ s, when the comoving horizon

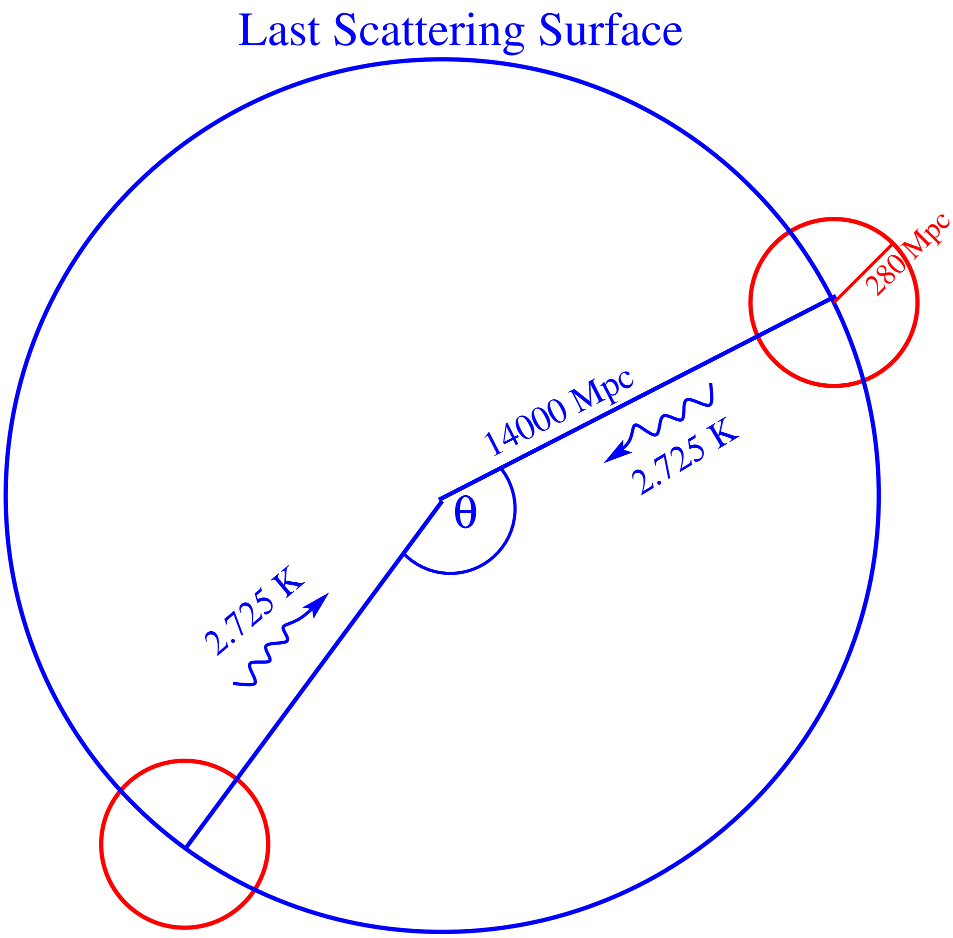

Figure 8.1: Horizon problem

size is

$$\eta_{\rm P} = \frac{a_{\rm P}}{H_0 {\Omega_{\rm r}}^{1/2}} \approx 5 \times 10^{-27}\ {\rm Mpc}, \tag{8.4}$$

there are

$$\left(\frac{\eta_0}{\eta_{\rm P}}\right)^3 = \left(\frac{14 \times 10^3}{5 \times 10^{-27}}\right)^3 \approx \left(10^{30}\right)^3 = 10^{90} \tag{8.5}$$

causally disconnected patches initially which must somehow have the exact same initial conditions. The sheer simplicity of the Universe thus causes a problem that is hard to explain.

8.1.2 Flatness problem

We also find that the curvature of the Universe is very small today, $\Omega_{\rm K} \ll 1$ or $\Omega_{\rm tot} = \sum_i \Omega_i \approx 1$, where i labels all particles that exist in the Universe including photons, neutrinos, baryons, dark matter, and dark energy, and Ω_i is the ratio of energy density in particular species to critical density today. We can however define a general quantity, $\Omega(t)$, at any time t using the Friedmann equation,

$$H^2 = \frac{8\pi G\rho(t)}{3} - \frac{K}{a(t)^2}. \tag{8.6}$$

At any time t, we can define in instantaneous ratio of total energy density in all particle species, $\rho(t)$ to the instantaneous critical density $\rho_{\rm cr}(t)$,

$$\Omega(t) = \frac{\rho(t)}{\rho_{\rm cr}(t)}, \tag{8.7}$$

where

$$\rho_{\rm cr}(t) = \frac{3H^2}{8\pi G} \tag{8.8}$$

is the energy density needed at time t for the Universe to be flat. We can therefore write the Friedmann equation as

$$\begin{aligned} H^2 &= \frac{8\pi G\rho_{\rm cr}(t)\Omega(t)}{3} - \frac{K}{a^2} \\ &= H^2\Omega - \frac{K}{a^2}. \end{aligned} \tag{8.9}$$

The curvature, $K = \pm 1, 0$ is a constant for any Universe. Thus,

$$H^2 a^2(\Omega - 1) = K = \text{constant}. \tag{8.10}$$

Using subscript $_{\rm ini}$ for values of quantitites at initial scale factor $a_{\rm ini}$, and subscript $_0$ for today, we have

$$\begin{aligned}\Omega_{\rm ini}-1 &= \frac{H_0^2 a_0^2}{H_{\rm ini}^2 a_{\rm ini}^2}\left(\Omega_0-1\right)\\ &= \frac{\dot{a}_0^2}{\dot{a}_{\rm ini}^2}\left(\Omega_0-1\right)\end{aligned} \tag{8.11}$$

For normal matter, such as dust and radiation, which constitutes the Universe at early times in simplest extrapolation of current Universe to $a \rightarrow 0$, the scale factor grows as a power law with respect to time, $a \propto t^n$ and $aH = \dot{a} \sim a/t \sim 1/\eta$. Specifically in radiation dominated era, $a \propto \eta$ and $aH \propto 1/a \propto 1/\eta$ is approximately the inverse of the comoving horizon. We therefore have,

$$\Omega_{\rm ini}-1 \sim \frac{\eta_{\rm ini}^2}{\eta_0^2}\left(\Omega_0-1\right) \sim 10^{-60}\left(\Omega_0-1\right), \tag{8.12}$$

for $\eta_{\rm ini} \sim \eta_{\rm p}$. Thus even if Ω_0 is not exactly unity but is of order unity today, i.e. Universe might be open or closed and not exactly flat today, initial Universe had to be incredibly close to flat. The Universe, during standard evolution in matter and radiation dominated eras become less and less flat with time. This is the flatness problem.

Another way to look at it is if $\Omega_{\rm ini}$ was just a little different (i.e. $10^{-60} \ll |\Omega_{\rm ini}-1| \ll 1$), the Universe would have expanded too fast or collapsed too quickly back to singularity to be consistent with the non-empty old Universe we see today. The initial energy density therefore had to be extermely finely tuned. In context of Newtonian dust Universe model, this corresponds to fine tuning the initial velocities of the particles to an incredible precision.

8.1.3 Initial fluctuations

We observe in the CMB small fluctuations of order $\Delta T/T \sim 10^{-4} - 10^{-5}$. These fluctuations are in fact observed to be present on super-horizon scales at the time of recombination. Moreover, analysis of CMB temperature and polarization anisotropies tells us that the initial perturbations are adiabatic and are present on super-horizon scales at very early times, not just at the time of recombination, giving rise to coherent acoustic oscillations i.e. Fourier modes which the same wavenumber magnitude, $k = |\mathbf{k}|$ all start to oscillate with the same phase $\propto \cos(kr_{\rm s})$. Also the amplitude of initial fluctuations is almost same on all scales i.e. they have a scale invariant power spectrum. The question now arises: how do you create correlations between causally disconnected patches of the Universe in a scale invariant way ? Causal sources of perturbations (such as cosmic strings) will give rise to incoherent evolution of perturbations since we could create a fluctuation on same scale k at different times which would start oscillating with different phases washing out the oscillation peaks of the CMB. The observation of coherent acoustic peaks in the CMB is evidence for super-horizon correlations of initial perturbations.

8.2 Motivation for inflation as solutions to problems of horizon, flatness and creation of initial perturbations

All of the above problems point to the fact that our observable Universe was somehow in in causal contact at early times. Therefore the naive extrapolation that the Universe remains dominated by radiation as $a \to 0$ is incorrect. One mechanism to restore causality at early times is *inflation*. In particular, the radiation dominated phase of the Universe cannot continue until the Planck scale but must be interrupted by an inflationary phase. The initial conditions at the beginning of the radiation dominated era, which are *the* initial conditions, $A(k)$, needed for the solutions in the radiation dominated and later phases of the Universe that we have been studying so far are *the end result* of a preceding inflationary phase. We will see that inflation does not completely solve the problem since the question about *the initial conditions at the beginning of inflation* is left open.

For normal matter, obeying the strong energy condition (SEC), the energy density ρ and pressure P satisfy

$$\rho + 3P \geq 0 \tag{8.13}$$

or the equation of state

$$w = \frac{P}{\rho} \geq -\frac{1}{3} \tag{8.14}$$

From Friedmann acceleration equation we see that

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \tag{8.15}$$

and Universe will always be decelerating as long as $w \geq -1/3$. This is in fact the physical content of the strong energy condition that gravity is always attractive, as far as normal matter is concerned. For non-relativistic matter $w \approx 0$ and for radiation $w = 1/3$. However it is possible that the strong energy condition is violated for some forms of matter, e.g. vaccum energy or scalar fields.

We can see how a violation of the SEC can help solve some of the problems of Friedmann cosmology by looking at the flatness problem. We see that in the Friedmann Universe the flatness problem arises because we always have deceleration and as a result $\dot{a}_{\rm ini} \gg \dot{a}_0$. However, if we had an accelerated expansion phase driven by matter with $w < -1/3$ in the early Universe, we could have $\dot{a}_{\rm ini} \sim \dot{a}_0$ or even $\dot{a}_{\rm ini} \ll \dot{a}_0$, and we avoid the fine tuning problem of making the Universe incredibly flat at very early times. In particular, we could start with arbitrary Ω and still produce the observable Universe with $\Omega \sim 1$.

The horizon problem is also solved. Lets take the case with $w \approx -1$ and $H \approx$ constant for simplicity i.e. early Universe dominated by matter behaving approximately as a cosmological constant (or Λ). The comoving horizon is given by

$$\begin{aligned}
\eta = \int_0^\eta {\rm d}\eta &= \int_{a_{\rm ini}}^{a_{\rm f}} \frac{{\rm d}\eta}{{\rm d}a}{\rm d}a \\
&= \int_{a_{\rm ini}}^{a_{\rm f}} \frac{{\rm d}t}{a{\rm d}a}{\rm d}a = \int_{a_{\rm ini}}^{a_{\rm f}} \frac{1}{a^2 H}{\rm d}a \\
&= \left.\frac{1}{aH}\right|_{a_f}^{a_{\rm ini}} = \frac{1}{H}\left[\frac{1}{a_{\rm ini}} - \frac{1}{a_{\rm f}}\right] \\
&\approx \frac{1}{a_{\rm ini}H}, \text{ for } a_{\rm f} \gg a_{\rm ini}
\end{aligned} \tag{8.16}$$

where $a_{\rm f}$ is the scale factor at the end of inflation. We therefore see that the comoving horizon at the end of inflation or beginning of the radiation dominated phase of the Universe is same as that at the beginning of inflation. It is much larger compared to the instantaneous horizon, $1/(a_{\rm f}H)$, at the end of inflation. The value $1/(a_{\rm f}H)$ is also approximately the horizon we would calculate if we had just extrpolated the radiation Universe backwards in time. It is a small initial causal patch which gives rise to the whole Universe today. This can be as large as we want if $a_{\rm ini}$ is small enough, i.e. if we start inflation early enough. Also we saw earlier that Λ dominated Universe has an event horizon,

$$r_{\rm EH} = \int_\eta^\infty {\rm d}\eta = \int_a^\infty \frac{a}{a^2 H}{\rm d}a = \frac{1}{aH} = \frac{1}{\dot{a}} \tag{8.17}$$

Since the Universe is accelerating, $\dot{a}$ is increasing and the comoving event horizon shrinks with time. Inhomogeneities outside the causal patch are therefore not able to influence and destroy the homogeneity inside the inflating patch.

8.2.1 Graceful exit

On differentiating the Hubble parameter w.r.t time we get

$$\dot{H} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{a}}{a}\right) = \frac{\ddot{a}}{a} - \frac{\dot{a}}{a^2}\dot{a} = \frac{\ddot{a}}{a} - H^2$$
$$\frac{\ddot{a}}{a} = \dot{H} + H^2. \tag{8.18}$$

During inflation $\ddot{a}/a > 0$. At the end of inflation we must transition to a radiation dominated decelerating Universe with $\ddot{a}/a < 0$. Therefore $\dot{H}$ should be negative, H should decrease with time and we should have $|\dot{H}| > H^2$ at some point so that the Universe can exit (*gracefully*) the inflationary expansion phase. Note that for a Universe dominated by a cosmological constant inflation never ends. The transition from accelerating to decelerating phase will happen when

$$\frac{|\dot{H}|}{H^2} \approx 1. \tag{8.19}$$

If we assume that $\dot{H}$ changes slowly compared to H, the duration of inflation is given by the time scale on which the initial H changes,

$$t_{\mathrm{inflation}} \sim \frac{H_{\mathrm{ini}}}{|\dot{H}_{\mathrm{ini}}|}. \tag{8.20}$$

We want the comoving horizon size at the beginning of inflation (i.e. size of the causal patch) to be of order of comoving horizon size today,

$$\frac{1}{a_{\mathrm{ini}}H_{\mathrm{ini}}} \approx \frac{1}{a_0 H_0}$$
$$\text{or} \quad \frac{\dot{a}_{\mathrm{ini}}}{\dot{a}_0} \approx 1 \tag{8.21}$$

We can impose an even stronger condition. Suppose we have perturbations of order unity in the initial Hubble patch of size $L_{\mathrm{ini}} \sim 1/(a_{\mathrm{ini}}H_{\mathrm{ini}})$,

$$\frac{\delta\rho_{\mathrm{ini}}}{\rho_{\mathrm{ini}}} \approx \frac{\nabla\rho_{\mathrm{ini}}L_{\mathrm{ini}}}{\rho_{\mathrm{ini}}} \approx \frac{\nabla\rho_{\mathrm{ini}}}{\rho_{\mathrm{ini}}a_{\mathrm{ini}}H_{\mathrm{ini}}} \sim 1, \tag{8.22}$$

where ∇ denotes the comoving derivative. Since the perturbations are frozen in during cosmological constant dominated expansion and the energy density is also approximately constant, $\nabla\rho/\rho$ is approximately constant during inflation and does not change substantially. At the end of inflation at time $t_{\mathrm{f}} \sim t_{\mathrm{inflation}}$, we have

$$\frac{\delta\rho_{\mathrm{f}}}{\rho_{\mathrm{f}}} \approx \frac{\nabla\rho_{\mathrm{f}}}{\rho_{\mathrm{f}}}L_{\mathrm{f}} \approx \frac{\nabla\rho_{\mathrm{ini}}}{\rho_{\mathrm{ini}}}\frac{1}{a_{\mathrm{f}}H_{\mathrm{f}}} = \left(\frac{\nabla\rho_{\mathrm{ini}}}{\rho_{\mathrm{ini}}}\frac{1}{a_{\mathrm{ini}}H_{\mathrm{ini}}}\right)\frac{a_{\mathrm{ini}}H_{\mathrm{ini}}}{a_{\mathrm{f}}H_{\mathrm{f}}}$$
$$\sim \frac{\dot{a}_{\mathrm{ini}}}{\dot{a}_{\mathrm{f}}}. \tag{8.23}$$

Since during inflation expansion is accelerating, $\dot{a}_{\rm f} \gg \dot{a}_{\rm ini}$ and any initial inhomogeneities are diluted by a factor of $\frac{\dot{a}_{\rm ini}}{\dot{a}_{\rm f}}$. In particular we want the inhomogeneities on the comoving horizon scale today, $1/(a_0 H_0) = 1/\dot{a}_0$ to be smaller than 10^{-5}. This means that we should have

$$\frac{\dot{a}_{\rm ini}}{\dot{a}_0} < 10^{-5}, \tag{8.24}$$

a much stronger condition compared to Eq. 8.21. The fluctuations on scales corresponding to the horizon size today must be small in the initial inflationary patch. In other words, it is not enough for the initial causal patch in which inflation happens to be of the size of observable Universe today, but we want the inflationary patch to be larger compared to the current observable Universe by a factor of 10^5,

$$10^{-5} > \frac{\dot{a}_{\rm ini}}{\dot{a}_0} = \frac{\dot{a}_{\rm ini}}{\dot{a}_{\rm f}}\frac{\dot{a}_{\rm f}}{\dot{a}_0} \approx \frac{a_{\rm ini}H_{\rm ini}}{a_{\rm f}H_{\rm f}}\frac{\dot{a}_{\rm f}}{\dot{a}_0} \approx \frac{a_{\rm ini}}{a_{\rm f}}\frac{\dot{a}_{\rm f}}{\dot{a}_0} \approx \frac{a_{\rm ini}}{a_{\rm f}}\frac{\eta_0}{\eta_{\rm P}} \approx 10^{30}\frac{a_{\rm ini}}{a_{\rm f}} \tag{8.25}$$

where we have used the result Eq. 8.5 assuming that inflation happens at the Planck energy scale and this is the energy scale at the end of inflation or beginning of the radiation dominated phase of the Universe. Note that the exact numbers will differ if inflation happened at a lower energy scale. We thus have

$$\frac{a_{\rm f}}{a_{\rm ini}} > 10^{35}\frac{H_{\rm ini}}{H_{\rm f}} \tag{8.26}$$

During inflation Hubble parameter is approximately constant ($w \approx -1$) and the scale factor grows exponentially,

$$\frac{a_{\rm f}}{a_{\rm ini}} = e^{H_{\rm ini}t_{\rm inflation}} > 10^{35}, \tag{8.27}$$

where $t_{\rm inflation}$ is the duration of inflation, Eq. 8.20. We therefore get the minimum duration inflation must last as

$$t_{\rm inflation} \approx \frac{H_{\rm ini}}{|\dot{H}_{\rm ini}|} > \ln(10^{35})H_{\rm ini}^{-1} \approx 80 H_{\rm ini}^{-1}. \tag{8.28}$$

Thus inflation must last longer than ~ 80 Hubble times[1]. For this to happen, Hubble must vary slowly,

$$\frac{|\dot{H}_{\rm ini}|}{H_{\rm ini}^2} < \frac{1}{80} \sim 10^{-2} \tag{8.29}$$

[1]Note that Mukhanov [74] obtains a value of 75, mostly because he uses for the current horizon size the value of proper time t_0

From Friedmann equations, (flat case, $K = 0$), we can derive

$$\dot{H} + H^2 = \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3P\right). \tag{8.30}$$

Rearranging terms, we get

$$\begin{aligned}
\frac{|\dot{H}|}{H^2} &= \frac{1}{H^2}\left|\frac{4\pi G}{3}(\rho + 3P) + H^2\right| \\
&= \frac{3}{8\pi G\rho}\left|\frac{4\pi G}{3}(\rho + 3P + 2\rho)\right| \\
&= \frac{3}{2}\left(\frac{\rho + P}{\rho}\right) \approx 10^{-2}
\end{aligned} \tag{8.31}$$

Therefore we get for the equation of state during inflation, w,

$$1 + w = \frac{\rho + P}{\rho} \approx 10^{-2} \tag{8.32}$$

The the equation of state during inflation should be just slightly different from -1, $w \approx -1 + 10^{-2}$, if we are to create our observable Universe from a patch initially in causal contact and then gracefully exit the inflationary expansion phase. This turns out to be a very generic and robust condition on inflationary models with important observational consequences [75].

8.3 Single field slow roll inflation

One way to realise a dynamical cosmological constant like behaviour is using a *classical* scalar field ϕ with Lagarangian density

$$\mathcal{L} = -\frac{1}{2}\left(\partial_\mu\phi\right)\partial^\mu\phi - V(\phi), \tag{8.33}$$

where subscript on ∂ denotes derivative w.r.t that space-time coordinate and $V(\phi)$ is the potential which determines the interactions of the field. We can calculate the stress energy tensor, $T^{\mu\nu}$, of the field ϕ using Noether's theorem, it is the conserved current corresponding to the invariance under space-time translations. Considered as an ideal fluid, $T^{\mu\nu}$ has only diagonal components given by (in Minkowski space-time with metric $\eta^{\mu\nu}$)

$$\begin{aligned}
T^{\mu\nu} &= -\frac{\partial\mathcal{L}}{\partial\left(\partial_\mu\phi\right)}\partial^\nu\phi + g^{\mu\nu}\mathcal{L} \\
&= \partial^\mu\phi\partial^\nu\phi - \eta^{\mu\nu}\left(\frac{1}{2}\partial_\alpha\phi\partial^\alpha\phi + V(\phi)\right).
\end{aligned} \tag{8.34}$$

Comparing with ideal fluid stress energy tensor,

$$T^{\mu\nu} = (\rho + P)\, U^{\mu} U^{\nu} + P\eta^{\mu\nu}, \tag{8.35}$$

where U^{μ} is the velocity of the observer, we get in the rest frame, $U^{\mu} = (1, 0, 0, 0)$, for the energy density and pressure of the scalar field ϕ,

$$\begin{aligned} \rho &= \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}\left(\nabla\phi\right)^2 + V(\phi) \\ P &= \frac{1}{3}\sum_i T^{ii}, \end{aligned} \tag{8.36}$$

where the index i denotes spatial components. For $i = 1$ we get

$$T^{11} = \left(\frac{\partial\phi}{\partial x^1}\right)^2 - \left(-\frac{1}{2}\dot{\phi}^2 + \frac{1}{2}\left(\nabla\phi\right)^2 + V(\phi)\right), \tag{8.37}$$

and similarly for the other two components. The pressure of scalar field is therefore

$$\begin{aligned} P = \frac{1}{3}\sum_i T^{ii} &= \frac{1}{3}\left(\nabla\phi\right)^2 + \frac{1}{2}\dot{\phi}^2 - \frac{1}{2}\left(\nabla\phi\right)^2 - V(\phi) \\ &= \frac{1}{2}\dot{\phi}^2 - V(\phi) - \frac{1}{6}\left(\nabla\phi\right)^2 \end{aligned} \tag{8.38}$$

If the field is homogeneous over the size of the horizon, we can neglect the spatial gradient terms, $\nabla\phi \sim 0$. In that case we have

$$\begin{aligned} \rho &\approx \frac{1}{2}\dot{\phi}^2 + V(\phi) \\ P &\approx \frac{1}{2}\dot{\phi}^2 - V(\phi) = -\rho + \dot{\phi}^2 \end{aligned} \tag{8.39}$$

Note that negelcting the spatial gradient, $\nabla\phi$, means that we are assuming the Universe is already homogeneous on scales comparable (or larger) compared to the horizon size initially. What is the probability for having such a patch in the early Universe is still an open question as we do not have a consistent theory on these energy scales which can provide a probability measure to calculate such probabilities. Arbitrary assumptions about the probability measure can lead to existence of such a patch, where necessary conditions for starting inflation are satisfied, to be extremely likely to exteremely unlikely.

If in addition the kinetic energy of the field is negligible compared to the potential energy, $\dot{\phi}^2 \ll V(\phi)$, we have $\rho \approx -P$, i.e. equation of state $w \approx -1$ and we therefore have accelerated expansion. In particular, we need to keep $\dot{\phi}^2 \ll V(\phi)$ for at least ~ 80 e-folds (or 80 Hubble times). As inflation proceeds, the kinetic energy grows and inflation ends when $\dot{\phi}^2 \approx V(\phi)$, $P \approx 0$ and the scalar field starts behaving like matter. We can get the equation of motion for the field from conservation of stress energy tensor. The continuity or energy conservation equation for an ideal fluid is

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + P). \tag{8.40}$$

We had derived this from Friedmann equations (Eq. 1.87), since conservation of stress-energy tensor is built into the equations of general relativity. For the scale field, ϕ, the continuity equation becomes

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0, \tag{8.41}$$

where a prime ($'$) denotes derivative w.r.t to ϕ, $V' \equiv \partial V/\partial\phi$. The Friedmann equation with the scalar field dominating the energy density of the Universe is

$$H^2 = \frac{8\pi}{3}\left(\frac{\dot{\phi}^2}{2} + V(\phi)\right), \tag{8.42}$$

where we have taken the Newton's gravitational constant $G = 1$, so that the energy is measured in units of the Planck energy given by Eq. 8.1.

To proceed further we must choose a potential $V(\phi)$. One of the simplest choices is

$$V(\phi) = \frac{1}{2}m^2\phi^2, \tag{8.43}$$

which gives the Klein-Gordon field. In quantized theory m is interpreted as the mass of the particle associated with the excitations of the Klein-Gordon field. Using the Friedmann equation, Eq. 8.42, in equation of motion for ϕ, Eq. 8.41, we get the equation for evolution of ϕ,

$$\ddot{\phi} + \sqrt{12\pi\left(\dot{\phi}^2 + m^2\phi^2\right)}\dot{\phi} + m^2\phi = 0 \tag{8.44}$$

There is no explicit time dependence in this equation. We can convert it to a first order differential equation by defining $y = \dot{\phi}$, where y is a function of ϕ, $y = y(\phi)$ and

$$\ddot{\phi} = \frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}\phi}\frac{\mathrm{d}\phi}{\mathrm{d}t} = y\frac{\mathrm{d}y}{\mathrm{d}\phi}. \tag{8.45}$$

With these substitutions, Eq. 8.44 can be written as

$$y\frac{\mathrm{d}y}{\mathrm{d}\phi} + \sqrt{12\pi\left(y^2 + m^2\phi^2\right)}y + m^2\phi = 0, \tag{8.46}$$

a first order ODE for $y(\phi)$. Rearranging we have

$$\frac{\mathrm{d}y}{\mathrm{d}\phi} = -\frac{\sqrt{12\pi\left(y^2 + m^2\phi^2\right)}y + m^2\phi}{y} \tag{8.47}$$

We can study the solutions of this equation using the trajectories of the field in the phase diagram in y, ϕ plane (see Fig. 8.2)

General expectation

The $V(\phi) = (1/2)m^2\phi^2$ potential has a minimum at $\phi = 0$ and quadratically rising with increasing $|\phi|$. The dynamics is similar to a ball rolling in a gravitational potential well. Without the Hubble friction term in Eq. 8.44, the field experiences an acceleration with magnitude $\propto |\phi|$ and directed towards $\phi = 0$. If ϕ started with a zero initial velocity, $\dot{\phi} = 0$, we would expect it to evolve towards the minimum at $\phi = 0$. There is a Hubble friction term in the acceleration equation of ϕ. This term serves to redshift away any initial velocities as the force $V'(\phi)$ accelerates the field toward $\phi = 0$. Once the initial velocities have redshifted away, in a suitable range of initial conditions, we can expect the Hubble friction H and $V'(\phi)$ to balance each other, reaching a steady state with $\dot{\phi}$ = constant. There is therefore an attractor solution or a steady state solution towards which any initial condition, within some range, might be expected to approach. This steady state solution will give us inflation if $\dot{\phi}^2 \ll m^2\phi^2$. To achieve this regime of small inflaton velocities, we need the accelerating force, $V'(\phi)$ to be small, i.e. the potential should be very flat. This is the *slow-roll* condition.

Inflation dynamics and attractor solutions

Lets now analyze the equations of motion of ϕ and find the conditions necessary for a successful inflation. In steady state the net acceleration vanishes, $\ddot{\phi} \approx 0$. This implies that

$$3H\dot{\phi} = -V'(\phi) = -m^2\phi. \tag{8.48}$$

Since we want the steady state solution to be the inflating solution with $\dot{\phi}^2 \ll V(\phi)$ and

$$\begin{aligned} H^2 &\approx \frac{8\pi}{3}V(\phi) = \frac{4\pi}{3}m^2\phi^2 \\ H &\approx \sqrt{\frac{4\pi}{3}}m|\phi|, \end{aligned} \tag{8.49}$$

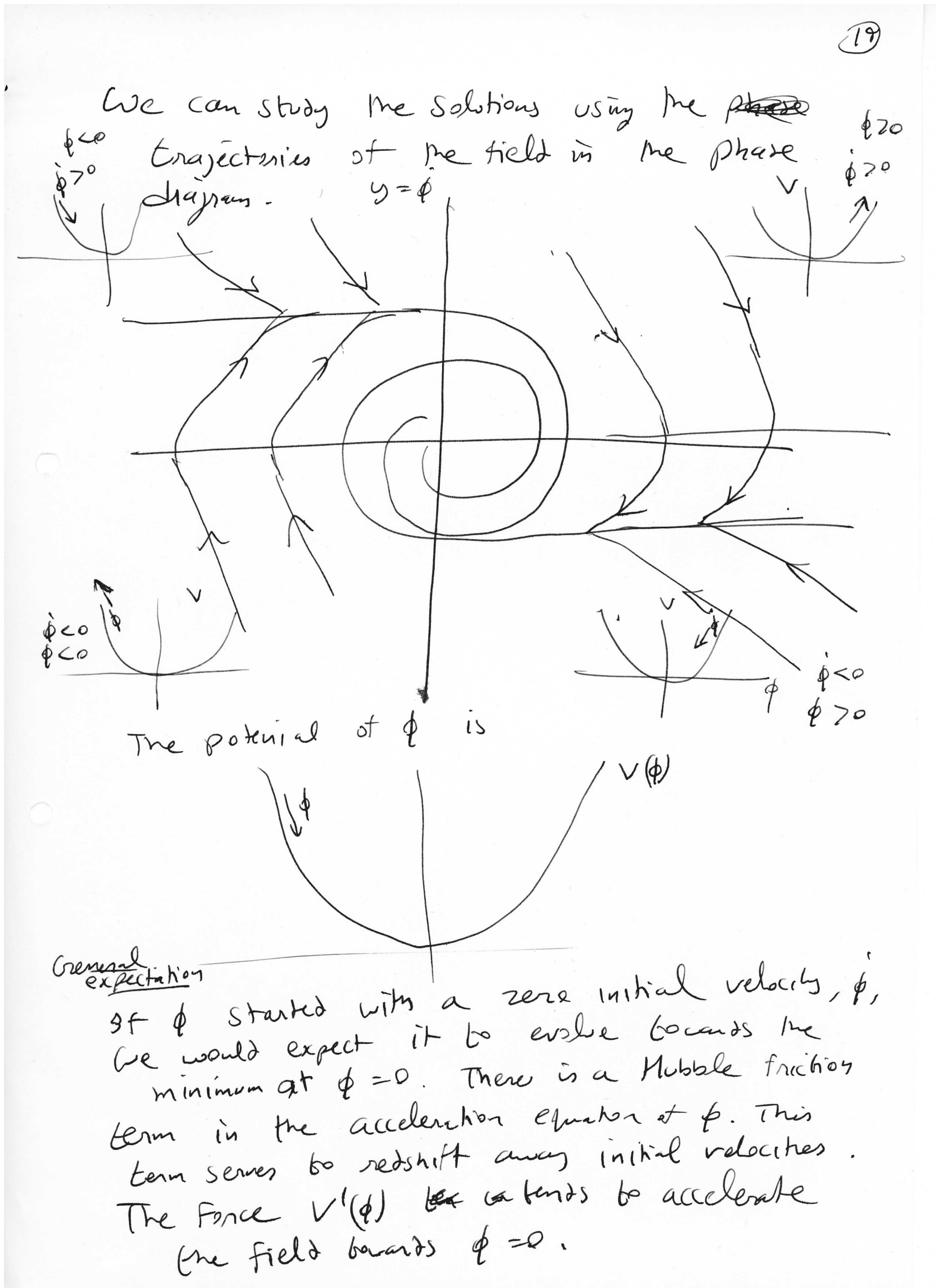

Figure 8.2: Inflaton trajectories

we get from Eq. 8.48

$$\dot{\phi} = -\frac{m^2\phi}{3H} = -\frac{m}{\sqrt{12\pi}}\mathrm{sign}(\phi), \tag{8.50}$$

whre sign(ϕ) = +1, −1 for $\phi > 0, < 0$ respectively. The condition $(1/2)\dot{\phi}^2 \ll V(\phi)$ in steady state is

$$\begin{aligned}\frac{1}{2}\frac{m^2}{12\pi} &\ll \frac{1}{2}m^2\phi^2 \\ \phi^2 &\gg \frac{1}{12\pi} \\ |\phi| &\gg \frac{1}{\sqrt{12\pi}} \approx 0.16 \end{aligned} \tag{8.51}$$

Therefore as $|\phi|$ approaches unity, the kinetic energy becomes important and we exit inflation. If we reach the steady state solution at $|\phi| \gg 1$, we should get enought inflation to create the observed Universe. The inflation end once $|\phi| \sim 1$, giving us a *graceful exit*.

Suppose we start away from the potential minimum with initial velocities, $|\phi| \gg 1$, $y^2 = \dot{\phi}^2 \gg V(\phi)$. In this case the evolution equation for y, Eq. 8.47, simplifies to

$$\frac{\mathrm{d}y}{\mathrm{d}\phi} = -\sqrt{12\pi y^2} = -\sqrt{12\pi}|y| \tag{8.52}$$

Lets analyze the solutions with initial conditions starting in each of the quadarants in the $y - \phi$ plane. For the initial conditions in the lower half plane, $y < 0$, $|y| = -y$ and the solution is given by

$$\begin{aligned}\ln y &= \sqrt{12\pi}\phi + \text{ constant} \\ \dot{\phi} = y &= Ce^{\sqrt{12\pi}\phi},\ C < 0, \end{aligned} \tag{8.53}$$

where C is the constant of intergration. Integrating again,

$$\frac{e^{-\sqrt{12\pi}\phi}}{-C\sqrt{12\pi}} = t + \text{ constant}. \tag{8.54}$$

We can make the constant above vanish by appropriately choosing the origin of time t. Solving for ϕ we get,

$$\phi = \text{ constant } - \frac{\ln t}{\sqrt{12\pi}}, \tag{8.55}$$

where the new constant includes C and factors of 12π. Since we are in the region of phase space with $\dot{\phi} < 0$, ϕ decreases with time. A small change in ϕ results in an exponential decrease in the magnitude of $\dot{\phi}$ (Eq. 8.53),

$$\dot{\phi} = -\frac{1}{\sqrt{12\pi}t}, \tag{8.56}$$

with $\dot{\phi}$ decreasing with time as $\propto 1/t$.

In the lower half plane, we have two possibilities with initial ϕ positive or negative. If we start at $\phi > 0$, the solution will hit the attractor in the upper right quadarant. If we start at $\phi < 0$, the solution will cross zero, $\dot{\phi} = 0$, and hit the other attractor. The equation of state during evolution towards the attractor us $\rho \approx P$. This is known as *stiff* equation of state with $w \approx +1$. The Friedmann equation in this regime is

$$\begin{aligned} H^2 &= \frac{8\pi}{6}\dot{\phi}^2 = \frac{1}{9t^2} \\ \frac{\dot{a}}{a} &= \frac{1}{3t} \end{aligned} \tag{8.57}$$

with solution

$$\begin{aligned} &a \propto t^{1/3} \\ &\rho \propto \dot{\phi}^2 \propto \frac{1}{t^2} \propto a^{-6}. \end{aligned} \tag{8.58}$$

The kinetic energy thus decreases (redshifts) faster with the expansion compraed to the case of radiation. At what value of ϕ we reach the attractor depends on the initial values of $\dot{\phi}$ and ϕ. Once on the attractor solution (Eq. 8.50, lower half plane attractor for $\phi > 0$)

$$\dot{\phi} = -\frac{m}{\sqrt{12\pi}}. \tag{8.59}$$

Integrating, we get

$$\begin{aligned} \phi(t) - \phi_{\rm i} &= \frac{m}{\sqrt{12\pi}}\left(-t + t_{\rm i}\right) \\ \phi(t) &= \phi_{\rm i} - \frac{m}{\sqrt{12\pi}}\left(t - t_{\rm i}\right). \end{aligned} \tag{8.60}$$

If we extrapolate this solution, from when we hit the attractor at $t = t_{\rm i}$ and start inflation to $\phi \approx 0$ when inflation ends at $t = t_{\rm f}$, we get the approximate duration of inflation. With the boundary condition that at $t = t_\mathrm{f}$, $\phi = 0$, we can rewrite Eq. 8.60 as

$$\begin{aligned}
\phi_\mathrm{i} &= \frac{m}{\sqrt{12\pi}}\left(t_\mathrm{f} - t_\mathrm{i}\right), \\
\phi(t) &= \frac{m}{\sqrt{12\pi}}\left(t_\mathrm{f} - t\right).
\end{aligned} \tag{8.61}$$

The Friedmann equation during the inflationary phase is

$$\begin{aligned}
H^2 &= \frac{8\pi}{3}V(\phi) = \frac{4\pi}{3}m^2\phi^2 \\
&= \frac{m^4}{9}\left(t_\mathrm{f} - t\right)^2 \\
\frac{\dot{a}}{a} &= \frac{m^2}{3}\left(t_\mathrm{f} - t\right)
\end{aligned} \tag{8.62}$$

We can solve this to get the evolution of the scale factor $a(t)$,

$$\begin{aligned}
\ln\left(\frac{a_\mathrm{f}}{a}\right) &= -\frac{m^2}{3}\left.\frac{(t_\mathrm{f} - t)^2}{2}\right|_t^{t_\mathrm{f}} = \frac{m^2}{6}(t_\mathrm{f} - t)^2 \\
a(t) &= a_\mathrm{f}e^{-m^2(t_\mathrm{f}-t)^2/6}
\end{aligned} \tag{8.63}$$

We can rewrite the solution in terms of the initial scale factor a_i at $t = t_i$ using

$$a_\mathrm{f} = a_\mathrm{i}e^{m^2(t_\mathrm{f}-t_\mathrm{i})^2/6} \tag{8.64}$$

to get

$$\begin{aligned}
a(t) &= a_\mathrm{i}e^{m^2\left[(t_\mathrm{f}-t_\mathrm{i})^2-(t_\mathrm{f}-t)^2\right]/6} \\
&= a_\mathrm{i}e^{m^2\left[(t_\mathrm{f}-t_\mathrm{i})(t_\mathrm{f}-t_\mathrm{i}+t_\mathrm{f}-t)\right]/6}
\end{aligned} \tag{8.65}$$

We can write the above solution in terms of the Hubble rate H and initial Hubble rate $H_\mathrm{i} = H(t_\mathrm{i})$

$$H(t) = \sqrt{\frac{4\pi}{3}}m\phi(t) = \frac{m^2}{3}\left(t_\mathrm{f} - t\right) \tag{8.66}$$

giving

$$a(t) = a_\mathrm{i}e^{\frac{1}{2}(t-t_\mathrm{i})(H+H_\mathrm{i})}, \tag{8.67}$$

since $H \propto \phi \propto t$, we have $\frac{1}{2}(H + H_\mathrm{i}) \approx H_\mathrm{avg}$, where H_avg is the average value of Hubble rate during inflation (i.e. during the steady state or attractor phase) and $a \propto e^{H_\mathrm{avg} t}$.

The change in the scale factor during inflation is

$$\frac{a_\mathrm{f}}{a_\mathrm{i}} = e^{(m^2/6)(t_\mathrm{f}-t_i)^2} = e^{2\pi|\phi_\mathrm{i}|^2}, \tag{8.68}$$

where ϕ_i and other initial variables labeled with subscript i refer to the value of these variables at the beginning of the inflationary case, at the instant when the solution hits one of the attractors and inflation begins. Thus the above equation just tells us the that earlier we start inflation (further from $\phi = 0$ minimum of potential), longer the inflation lasts.

For a field of mass m, if we want to remain in the classical gravity or sub-planckian regime, the energy density should remain below the Planck scale. Since energy density is dominated by the potential energy, we have the condition,

$$\begin{aligned} V(\phi) \approx m^2\phi^2 \lesssim E_\mathrm{P}^4 = \left(10^{19}\ \mathrm{GeV}\right)^4 \\ \phi^2 \lesssim \frac{\left(10^{19}\ \mathrm{GeV}\right)^4}{m^2}. \end{aligned} \tag{8.69}$$

The upper limit for the field value, and hence the maximum amount of inflation we can have, depends on the mass parameter m of the field. For example, if $m = 10^{13}$ Gev,

$$\begin{aligned} \phi^2 &\lesssim 10^{12} E_\mathrm{P}^2 \\ |\phi| &\lesssim 10^6 E_\mathrm{P} = 10^6 \end{aligned} \tag{8.70}$$

in Planck units and the maximum allowed inflation is

$$\frac{a_\mathrm{f}}{a_\mathrm{i}} \sim e^{|\phi_\mathrm{i}|^2} \sim e^{10^{12}} \tag{8.71}$$

or 10^{12} e-folds. Remember that we needed a *minimum* of ~ 80 e-folds to create the observable Universe from reasonable initial conditions at Planck scale. Therefore we should expect that for a large volume of parameter space in the initial $\phi_\mathrm{i} - \dot{\phi}_\mathrm{i}$ plane, we can get the required inflation. Thus there is no need to fine tune the initial values for the field ϕ. We still need the field to be uniform over few Hubble volumes since we neglected the gradient terms, $\nabla\phi$.

Once the field reached $\phi = 0$, it will not stop there but overshoot, since it will have a finite velocity, $\dot{\phi}$, and start to oscillate about the minimum. In analogy with an oscillating pendulum, the average kinetic energy of the field over an oscillation would be equal to the average potential energy.

We can find the solution along the attractor in the limit of small ϕ to study the dynamics at the end of inflation, near $\phi = 0$. Anticipating an oscillating solution, we can do a change of variables,

$$\begin{aligned} \dot{\phi} &= A\cos\theta \\ m\phi &= A\sin\theta, \end{aligned} \tag{8.72}$$

so that the average kinteic energy over an oscillation = $\langle \frac{1}{2}\dot{\phi}^2\rangle = \langle \frac{1}{2}m^2\phi^2 = \frac{1}{2}A^2\langle\cos^2\theta\rangle = \frac{1}{4}A^2$ = average potential energy. In this limit, the average pressure is given by

$$\langle P\rangle = \frac{1}{2}\dot{\phi}^2\rangle - \langle\frac{1}{2}m^2\phi^2 = 0, \tag{8.73}$$

i.e. the field will start behaving like cold matter at zero temperature. The Friedmann equation gives us

$$\begin{aligned} H^2 &= \frac{8\pi}{3}\left(\frac{1}{2}\dot{\phi}^2 + \frac{1}{2}m^2\phi^2\right) = \frac{4\pi}{3}A^2 \\ A &= \sqrt{\frac{3}{4\pi}}H. \end{aligned} \tag{8.74}$$

The equation of motion for field ϕ, Eq. 8.44, in terms of the new variables is given by

$$\dot{A}\cos\theta - A\dot{\theta}\sin\theta + 3HA\cos\theta + A\sin\theta = 0. \tag{8.75}$$

Using Eq. 8.72 we can write

$$m\dot{\phi} = \dot{A}\sin\theta + A\dot{\theta}\cos\theta = mA\cos\theta. \tag{8.76}$$

Multiplying Eq. 8.75 by $\sin\theta$ and Eq. 8.76 by $\cos\theta$ and subtracting we get

$$\begin{aligned} A\dot{\theta} - 3AH\cos\theta\sin\theta - mA &= 0 \\ \dot{\theta} = m + 3H\cos\theta\sin\theta &= m + \frac{3}{2}H\sin(2\theta). \end{aligned} \tag{8.77}$$

Multiplying Eq. 8.75 by $\cos\theta$ and Eq. 8.76 by $\sin\theta$ and adding we get

$$\dot{A} + 3HA\cos^2\theta = 0 \tag{8.78}$$

and using Eq. 8.74 we have

$$\dot{H} = -3H^2\cos^2\theta. \tag{8.79}$$

The Hubble parameter therefore continues to decrease in the matter-like phase as expected. Once $H \ll m$, we can neglect the second term in Eq. 8.77 giving

$$\begin{aligned}\dot{\theta} &= m \\ \theta &= mt + \text{ constant}\end{aligned} \tag{8.80}$$

The constant above is just the phase of oscillation which we can set to zero by shifting the time t by an appropriate amount. Substituting the solution for θ in Eq. 8.79 we get

$$\dot{H} = -3H^2 \cos^2(mt). \tag{8.81}$$

Integrating we get

$$\begin{aligned}\int \frac{\mathrm{d}H}{H^2} &= -3 \int \cos^2(mt)\mathrm{d}t = -\frac{3}{2} \int (1 + \cos(2mt))\, \mathrm{d}t \\ -\frac{1}{H} &= -\frac{3}{2t}\left(1 - \frac{\sin(2mt)}{2mt}\right) \\ H &= \frac{2}{3t}\left(1 - \frac{\sin(2mt)}{2mt}\right)^{-1},\end{aligned} \tag{8.82}$$

where the constant of integration can be again made negligible by translating time. This solution is applicable, as stated above, when $H \ll m$ or $t \gg 1/m$. Since during the oscillating phase, on average, $\langle \cos^2 \theta \rangle = 1/2$, from Eq. 8.81, on average,

$$\begin{aligned}\dot{H} &= -\frac{3}{2}H^2 \\ H &\propto \frac{1}{t}.\end{aligned} \tag{8.83}$$

At $mt \gg 1$, we can Taylor expand the solution, Eq. 8.81,

$$H = \frac{\dot{a}}{a} \approx \frac{2}{3t}\left(1 + \frac{\sin(2mt)}{2mt}\right) + O\left(\frac{m}{(mt)^3}\right) \tag{8.84}$$

Integrating again to solve for the scale factor,

$$\begin{aligned}\ln a &= \frac{2}{3} \ln t + \frac{2}{3}\left[\mathrm{Ci}(2mt) - \frac{\sin(2mt)}{2mt}\right] + O\left(\frac{1}{(mt)^3}\right) \\ &\approx \frac{2}{3} \ln t + \frac{2}{3}\left[\frac{-\cos(2mt)}{(2mt)^2}\right]\end{aligned} \tag{8.85}$$

where we have used the asymptotic expansion for the cosine integral as $x \to \infty$,

$$\mathrm{Ci}(x) = -\int_x^\infty \frac{\cos t}{t}\mathrm{d}t \approx \frac{\sin(x)}{x}\left(1 - \frac{2}{x^2} + \cdots\right) - \frac{\cos(x)}{x^2}\left(1 - \frac{6}{x^2} + \cdots\right). \quad (8.86)$$

We thus get the solution for scale factor,

$$a \approx t^{2/3} e^{-\frac{\cos(2mt)}{6m^2t^2}} \approx t^{2/3}\left(1 - \frac{\cos(2mt)}{6m^2t^2}\right) \quad (8.87)$$

Thus the scale factor evolves as $\propto t^{2/3}$, similar to the matter dominated era, with small oscillatory corrections which damp over time. The solution for the field ϕ is given by (Eq. 8.72)

$$\begin{aligned}\phi(t) &= \frac{A}{m}\sin\theta = \sqrt{\frac{3}{4\pi}}\frac{H}{m}\sin(mt) \\ &\approx \sqrt{\frac{1}{3\pi}}\frac{\sin(mt)}{mt}\left(1 + \frac{\sin(2mt)}{2mt}\right)\end{aligned} \quad (8.88)$$

Thus the field oscillates with ever decreasing amplitude, $\phi^2 \propto t^{-2} \propto a^{-3}$, as is expected from a damped oscillator. The friction is provided by the Hubble expansion which redshifts the energy density of the field as in matter dominated era.

The oscillating field, with the same average equation of state as cold matter, is equivalent to a condensate of massive ϕ particles in zero momentum (ground) state. Note that no such particle interpretation is possible of the classical field in the inflationary phase, where we do not know of any particles which would behave with $w = -1$ or $w = +1$. A useful analogy is with the electromagnetic fields. Static electric or magnetic field does not allow a particle description (e.g. field between plates of a capacitor). Changing or oscillating electromagnetic field, such as radio waves, allows a description in terms of coherent superposition of large number of real particles, photons each with energy $\hbar\omega$, where ω is the oscillating frequency of the electromagnetic wave.

The decay of these massive scalar particles at the end of inflation, either directly or through an intermediate particle into the standard model particles begins the standard radiation dominated phase of the Universe.

8.3.1 Inflation with general potentials

At the end of inflation we are already in a decelerating Friedmann Universe. Thus we can smoothly transition from an accelerating phase to a decelerating phase by an appropriate choice of dynamics of a scalar field. Also note that the oscillating solution, and the real massive inflaton particles, appear automatically at the end of inflation and is the result of a quadratic potential of the field. The $m^2\phi^2$ potential results in non-interacting ϕ particles of mass m on quantization. In general the particles may have interactions which would result in ϕ^n, $n > 2$ terms in the potential. For a general potential, $V(\phi)$, we want to derive conditions which allow inflationary solutions. The general features of the solutions we discussed for $m^2\phi^2$ potentials apply to the general potential also. Friection would still damp initial velocities until ϕ reaches *terminal* velocities defined by only the $V(\phi)$. If such a steady state solution is achieved, then we will have $\ddot{\phi} \approx 0$ giving *with $\dot{\phi}^2 \ll V(\phi)$) for the equation of motion of ϕ and Friedmann equation,

$$3H\dot{\phi} + V'(\phi) \approx 0 \tag{8.89}$$

$$H \approx \sqrt{\frac{8\pi}{3}V(\phi)}. \tag{8.90}$$

To be specific, we want the acceleration to be small compared to either of the force terms in the equation of motion of ϕ, i.e. we have the condition $\ddot{\phi} \ll 3H\dot{\phi}$ in steady state which can be written as

$$1 \gg \left|\frac{\ddot{\phi}}{3H\dot{\phi}}\right| \approx \left|\frac{\ddot{\phi}}{V'(\phi)}\right| \tag{8.91}$$

Taking the time derivatve of Eq. 8.89 we have

$$\begin{aligned} 3H\ddot{\phi} + 3\dot{H}\dot{\phi} + V''\dot{\phi} &\approx 0 \\ \left|\frac{\ddot{\phi}}{\dot{\phi}3H}\right| &\approx \left|\frac{-3\dot{H} - V''}{9H^2}\right| \\ &\approx \left|\frac{-3\dot{H} - V''}{\frac{8\pi}{3}V}\right| \\ &\sim \left|\frac{V''}{V}\right| \ll 1, \end{aligned} \tag{8.92}$$

if $|\dot{H}| \ll |V|$. We can check that this is satisfied as follows. The condition $|\dot{\phi}^2| \ll |V|$ gives

$$\left|\frac{\dot{\phi}^2}{V}\right| \approx \left|\frac{(V')^2}{9H^2V}\right| \sim \left|\frac{V'}{V}\right|^2 \ll 1 \tag{8.93}$$

Taking derivative of the Friedmann equation gives

$$2H\dot{H} \approx V'\dot{\phi} \approx V'\frac{V'}{3H}. \tag{8.94}$$

We therefore have

$$\frac{\dot{H}}{V} \approx \frac{(V')^2}{VH^2} \approx \left(\frac{V'}{V}\right)^2 \ll 1 \tag{8.95}$$

The conditions Eq. 8.92 and 8.93 are called slow roll conditions and constrain how fast the potential can vary as a function of ϕ so that we can have inflation. For general power law potentials,

$$\begin{aligned} V &= \frac{1}{n}\lambda\phi^n \\ V' &= \lambda\phi^{n-1} \\ V'' &= \lambda(n-1)\phi^{n-2} \end{aligned} \tag{8.96}$$

the slow roll conditions are

$$\begin{aligned} \left|\frac{V'}{V}\right|^2 &\approx \frac{1}{\phi^2} \\ \frac{V''}{V} &\approx \frac{1}{\phi^2} \end{aligned} \tag{8.97}$$

Thus for $|\phi| \gg 1$ (in Planck units), slow roll conditions are satisfied for power law potentials. For general power law (n), we can write equation of motion, neglecting expansion, at the end of inflation in the oscillatory phase as,

$$\begin{aligned} \ddot{\phi} + V' &= 0 \\ \frac{\mathrm{d}\left(\phi\dot{\phi}\right)}{\mathrm{d}t} - \dot{\phi}^2 + \phi V' &= 0 \end{aligned} \tag{8.98}$$

Averaging over an oscillation, $\langle\phi\dot{\phi}\rangle \approx 0$. We therefore have

$$\dot{\phi}^2 \approx \phi V' = \phi\lambda\phi^{n-1} = \lambda\phi^n = nV \tag{8.99}$$

The equation of state at the end of inflation is therefore

$$w = \frac{P}{\rho} = \frac{\frac{1}{2}\dot{\phi}^2 - V}{\frac{1}{2}\dot{\phi}^2 + V} \approx \frac{nV - 2V}{nV + 2V} = \frac{n-2}{n+2}. \tag{8.100}$$

For $V = m^2\phi^2$ we get $w \approx 0$ and for $n = 4$, we get $w = 1/3$. Thus the particles at the end of inflation are ultrarelativistic for ϕ^4 potential.

Additional references

A very good reference for the particle production in *new inflationary cosmology* is [76]. For a discussion on the cosmological constant problem see [77].

Bibliography

[1] Thomas Wright. An original Theory or New Hypothesis of the Universe, 1750.

[2] F. A. Paneth. Thomas Wright of Durham and Immanuel Kant. *The Observatory*, 64:71–82, Jun 1941. [ADS].

[3] Immanuel Kant. Allgemeine Naturgeschichte und Theorie des Himmels, Part I, 1755.

[4] Kurtiss J. Gordon. Messier 33, History of our Understanding of a Spiral Galaxy. *Quarterly Journal of the Royal Astronomical Society*, 10:293, Dec 1969. [ADS].

[5] Harlow Shapley and Heber D. Curtis. The Scale of the Universe. *Bulletin of the National Research Council*, 2(11):171–217, May 1921. [ADS].

[6] V. M. Slipher. Spectrographic Observations of Nebulae. *Popular Astronomy*, 23:21–24, Jan 1915. [ADS].

[7] Henrietta S. Leavitt and Edward C. Pickering. Periods of 25 Variable Stars in the Small Magellanic Cloud. *Harvard College Observatory Circular*, 173:1–3, Mar 1912. [ADS].

[8] E. P. Hubble. Cepheids in spiral nebulae. *The Observatory*, 48:139–142, May 1925. [ADS].

[9] E. P. Hubble. A spiral nebula as a stellar system: Messier 33. *ApJ*, 63:236–274, May 1926. [DOI], [ADS].

[10] Virginia Trimble. The 1920 Shapley-Curtis Discussion: Background, Issues, and Aftermath. *PASP*, 107:1133, Dec 1995. [DOI], [ADS].

[11] Yaroslav A. Lazovik, Alexey S. Rastorguev, Marina V. Zabolotskikh, and Natalia A. Gorynya. The period-luminosity relation for Cepheids derived

from multiphase temperature measurements and Cepheids kinematics based on GAIA DR2 data. *arXiv e-prints*, page arXiv:1911.13102, Nov 2019. `arXiv:1911.13102`, [ADS].

[12] E. E. Mamajek, G. Torres, A. Prsa, P. Harmanec, M. Asplund, P. D. Bennett, N. Capitaine, J. Christensen-Dalsgaard, E. Depagne, W. M. Folkner, M. Haberreiter, S. Hekker, J. L. Hilton, V. Kostov, D. W. Kurtz, J. Laskar, B. D. Mason, E. F. Milone, M. M. Montgomery, M. T. Richards, J. Schou, and S. G. Stewart. IAU 2015 Resolution B2 on Recommended Zero Points for the Absolute and Apparent Bolometric Magnitude Scales. *arXiv e-prints*, page arXiv:1510.06262, Oct 2015. `arXiv:1510.06262`, [ADS].

[13] M. M. Phillips. The Absolute Magnitudes of Type IA Supernovae. *ApJL*, 413:L105, Aug 1993. `[DOI]`, [ADS].

[14] Edwin Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173, Mar 1929. `[DOI]`, [ADS].

[15] A. Friedmann. Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386, 1922. `[DOI]`, [ADS].

[16] A. Friedmann. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift fur Physik*, 21:326–332, December 1924. `[DOI]`, [ADS].

[17] G. Lemaître. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Societe Scietifique de Bruxelles*, 47:49–59, 1927. [ADS].

[18] V. Trimble. Anybody but Hubble! *ArXiv e-prints*, July 2013. `arXiv:1307.2289`, [ADS].

[19] Licia Verde, Tommaso Treu, and Adam G. Riess. Tensions between the early and late Universe. *Nature Astronomy*, 3:891–895, Sep 2019. `arXiv:1907.10625`, `[DOI]`, [ADS].

[20] Gianfranco Bertone and Dan Hooper. History of dark matter. *Reviews of Modern Physics*, 90(4):045002, Oct 2018. `arXiv:1605.04909`, `[DOI]`, [ADS].

[21] H. P. Robertson. On the Foundations of Relativistic Cosmology. *Proceedings of the National Academy of Science*, 15:822–829, November 1929. `[DOI]`, [ADS].

[22] H. P. Robertson. Kinematics and World-Structure. *ApJ*, 82:284, November 1935. [DOI], [ADS].

[23] A. G. Walker. On milne's theory of world-structure. *Proceedings of the London Mathematical Society*, s2-42(1):90–127, 1937. [DOI].

[24] Suman Sarkar and Biswajit Pandey. An information theory based search for homogeneity on the largest accessible scale. *MNRAS*, 463(1):L12–L16, Nov 2016. arXiv:1607.06194, [DOI], [ADS].

[25] S. Weinberg. *Cosmology*. Oxford University Press, Oxford, 2008.

[26] A. D. Dolgov. Neutrinos in cosmology. *Physics Reports*, 370:333–535, November 2002. arXiv:hep-ph/0202122, [DOI], [ADS].

[27] R. H. Cyburt, B. D. Fields, K. A. Olive, and T.-H. Yeh. Big bang nucleosynthesis: Present status. *Reviews of Modern Physics*, 88(1):015004, January 2016. arXiv:1505.01076, [DOI], [ADS].

[28] Brian D. Fields, Paolo Molaro, and Subir Sarkar. Big-Bang Nucleosynthesis. *arXiv e-prints*, page arXiv:1412.1408, Dec 2014. arXiv:1412.1408, [ADS].

[29] M. Tanabashi et al. and Particle Data Group. Review of Particle Physics*. *Phys.Rev.D*, 98(3):030001, Aug 2018. [DOI], [ADS].

[30] G. Jungman, M. Kamionkowski, and K. Griest. Supersymmetric dark matter. *Physics Reports*, 267:195–373, 1996. [DOI], [ADS].

[31] E. Aprile et al. and Xenon Collaboration. Dark Matter Search Results from a One Ton-Year Exposure of XENON1T. *Phys.Rev.Lett*, 121(11):111302, Sep 2018. arXiv:1805.12562, [DOI], [ADS].

[32] D. Pequignot, P. Petitjean, and C. Boisson. Total and effective radiative recombination coefficients. *A&A*, 251:680–688, November 1991. [ADS].

[33] Y. Ali-Haïmoud and C. M. Hirata. Ultrafast effective multilevel atom method for primordial hydrogen recombination. *Phys.Rev.D*, 82(6):063521, 2010. [DOI], [ADS].

[34] Y. Ali-Haïmoud and C. M. Hirata. HyRec: A fast and highly accurate primordial hydrogen and helium recombination code. *Phys.Rev.D*, 83(4):043513, 2011. [DOI], [ADS].

[35] J. Chluba and R. A. Sunyaev. Free-bound emission from cosmological hydrogen recombination. *A&A*, 458:L29–L32, 2006. [DOI], [ADS].

[36] Y. B. Zeldovich, V. G. Kurt, and R. A. Sunyaev. Recombination of Hydrogen in the Hot Model of the Universe. *Zh. Eksp. Teor. Fiz.*, 55:278, 1968. [ADS].

[37] P. J. E. Peebles. Recombination of the Primeval Plasma. *ApJ*, 153:1, 1968. [ADS].

[38] V. V. Sobolev. *Moving envelopes of stars*. Harvard University Press, Cambridge, 1960. [ADS].

[39] S. Seager, D. D. Sasselov, and D. Scott. A New Calculation of the Recombination Epoch. *ApJL*, 523:L1–L5, September 1999. arXiv:astro-ph/9909275, [DOI], [ADS].

[40] J. Chluba and R. M. Thomas. Towards a complete treatment of the cosmological recombination problem. *MNRAS*, 412:748–764, 2011. URL: http://www.chluba.de/cosmorec, [DOI], [ADS].

[41] J. Chluba and Y. Ali-Haïmoud. COSMOSPEC: fast and detailed computation of the cosmological recombination radiation from hydrogen and helium. *MNRAS*, 456:3494–3508, March 2016. arXiv:1510.03877, [DOI], [ADS].

[42] J. Binney and S. Tremaine. *Galactic Dynamics*. Princeton University Press, Princeton, 2008.

[43] J. H. Jeans. The Stability of a Spherical Nebula. *Philosophical Transactions of the Royal Society of London Series A*, 199:1–53, January 1902. [DOI], [ADS].

[44] E. M. Lifshitz. On the gravitational stability of the expanding universe. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, 16:587–602, 1946. [ADS].

[45] W. B. Bonnor. Jeans' formula for gravitational instability. *MNRAS*, 117:104, January 1957. [DOI], [ADS].

[46] Hideo Kodama and Misao Sasaki. Cosmological Perturbation Theory. *Progress of Theoretical Physics Supplement*, 78:1, January 1984. [DOI], [ADS].

[47] S. Dodelson. *Modern cosmology*. Academic Press, Amsterdam, 2003. [ADS].

[48] A. Kosowsky. Cosmic microwave background polarization. *Annals of Physics*, 246:49–85, February 1996. arXiv:astro-ph/9501045, [DOI], [ADS].

[49] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii. *Quantum Theory of Angular Momentum*. World Scientific, Singapore, 1988. [DOI], [ADS].

[50] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al. Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13, September 2016. arXiv:1502.01589, [DOI], [ADS].

[51] R. Keisler et al. A Measurement of the Damping Tail of the Cosmic Microwave Background Power Spectrum with the South Pole Telescope. *ApJ*, 743:28, 2011. [DOI], [ADS].

[52] R. Hlozek et al. The Atacama Cosmology Telescope: A Measurement of the Primordial Power Spectrum. *ApJ*, 749:90, 2012. [DOI], [ADS].

[53] W. Hu and N. Sugiyama. Anisotropies in the cosmic microwave background: an analytic approach. *ApJ*, 444:489–506, May 1995. arXiv:astro-ph/9407093, [DOI], [ADS].

[54] Wayne Hu, Naoshi Sugiyama, and Joseph Silk. The physics of microwave background anisotropies. *Nature*, 386(6620):37–43, March 1997. arXiv:astro-ph/9504057, [DOI], [ADS].

[55] L. D. Landau and E. M. Lifshitz. *Fluid Mechanics*. Butterworth-Heinemann, Oxford, 1987.

[56] R. Khatri, R. A. Sunyaev, and J. Chluba. Mixing of blackbodies: entropy production and dissipation of sound waves in the early Universe. *A&A*, 543:A136, 2012. [DOI], [ADS].

[57] Scott Dodelson. The Real Problem with MOND. *International Journal of Modern Physics D*, 20(14):2749–2753, January 2011. arXiv:1112.1320, [DOI], [ADS].

[58] Sergei Bashinsky and Uroš Seljak. Signatures of relativistic neutrinos in CMB anisotropy and matter clustering. *Phys.Rev.D*, 69(8):083002, April 2004. arXiv:astro-ph/0310198, [DOI], [ADS].

[59] J. Silk. Cosmic Black-Body Radiation and Galaxy Formation. *ApJ*, 151:459, 1968. [DOI], [ADS].

[60] P. J. E. Peebles and J. T. Yu. Primeval Adiabatic Perturbation in an Expanding Universe. *ApJ*, 162:815, 1970. [DOI], [ADS].

[61] N. Kaiser. Small-angle anisotropy of the microwave background radiation in the adiabatic theory. *MNRAS*, 202:1169–1180, 1983. [ADS].

[62] R. A. Sunyaev and Y. B. Zeldovich. The interaction of matter and radiation in the hot model of the Universe, II. *ApSS*, 7:20–30, 1970. [DOI], [ADS].

[63] L. Danese and G. de Zotti. Double Compton process and the spectrum of the microwave background. *A&A*, 107:39–42, 1982. [ADS].

[64] J. Chluba and R. A. Sunyaev. The evolution of CMB spectral distortions in the early Universe. *MNRAS*, 419:1294–1314, 2012. [DOI], [ADS].

[65] R. Khatri and R. A. Sunyaev. Creation of the CMB spectrum: precise analytic solutions for the blackbody photosphere. *JCAP*, 6:38, 2012. [DOI], [ADS].

[66] Y. B. Zeldovich and R. A. Sunyaev. The Interaction of Matter and Radiation in a Hot-Model Universe. *ApSS*, 4:301–316, 1969. [DOI], [ADS].

[67] A. S Kompaneets. The establishment of thermal equilibrium between quanta and electrons. *Zh. Eksp. Teor. Fiz.*, 31:876–875, 1956.

[68] R. K. Sachs and A. M. Wolfe. Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. *ApJ*, 147:73, January 1967. [DOI], [ADS].

[69] Sergei Bashinsky. Gravity of cosmological perturbations in the CMB. *Phys.Rev.D*, 74(4):043007, August 2006. arXiv:astro-ph/0405157, [DOI], [ADS].

[70] U. Seljak and M. Zaldarriaga. A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *ApJ*, 469:437, 1996. [DOI], [ADS].

[71] M. J. Rees and D. W. Sciama. Large-scale Density Inhomogeneities in the Universe. *Nature*, 217(5128):511–516, February 1968. [DOI], [ADS].

[72] Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, page arXiv:1807.06209, July 2018. arXiv:1807.06209, [ADS].

[73] Eric Hivon, Krzysztof M. Górski, C. Barth Netterfield, Brendan P. Crill, Simon Prunet, and Frode Hansen. MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets. *ApJ*, 567(1):2–17, March 2002. arXiv:astro-ph/0105302, [DOI], [ADS].

[74] V. Mukhanov. *Physical Foundations of Cosmology*. Cambridge University Press, Cambridge, November 2005. [DOI], [ADS].

[75] V. F. Mukhanov and G. V. Chibisov. Quantum fluctuations and a nonsingular universe. *Soviet Journal of Experimental and Theoretical Physics Letters*, 33:532, May 1981. [ADS].

[76] L. F. Abbott, Edward Farhi, and Mark B. Wise. Particle production in the new inflationary cosmology. *Physics Letters B*, 117(1-2):29–33, November 1982. [DOI], [ADS].

[77] E. W. Kolb and S. Wolfram. Spontaneous symmetry breaking and the expansion rate of the early universe. *ApJ*, 239:428–432, July 1980. [DOI], [ADS].