Relativity and Cosmology II

Erwin Tanin and Adrien Florio

Based on Prof. Mikhail Shaposhnikovś lectures in Spring 2018

Introduction

You have probably stumbled upon some of the following statements about the universe

- The universe is homogeneous and isotropic.
- The universe is currently expanding at an accelerated rate.
- The universe was extremely hot and dense in the past.
- The universe contains relic black-body radiation.
- The universe contains two types of matter: ordinary matter and dark matter.
- The universe contains dark energy.
- Light elements such as D, ^{3}He, ^{4}He, and Li in the universe were cooked during the first few minutes after the Big Bang.
- The universe was inflating in the past.
- Large-scale structures of the universe (clusters of galaxies) are originated from quantum fluctuations.

The aim of these lectures is to explain the above statements with the use of general relativity, equilibrium and non-equilibrium statistical mechanics, and particle physics. The notations used here follow from that of *Relativity and Cosmology I*.

In this note, the main lecture material is interspersed with occasional supplementary materials: blocks of text of smaller font-size which can be skipped during the first reading. These supplementary materials provide side remarks or touch upon more advanced topics that are beyond the scope of these lectures. At the end of each lecture, exercises of various difficulty levels are provided. Some of these exercises are solved explicitly, in which case they are called examples. This note is also equipped with a short bibliography at the end of each part, referring mostly to textbooks that explain a particular topic in more detail and occasionally to articles deriving specific results that are not derived here. A list of theory questions is given at the end of the note.

Contents

Expanding Universe

Lecture 1

Large-scale properties of the universe; FRW metric.

1.1 Isotropy and Homogeneity of the Universe

In order to give a sense of scales for various objects studied in cosmology and astrophysics, we list here the sizes of some of the most recognizable structures in the universe:

- Earth radius $\approx 6.4 \times 10^8$ cm
- Solar radius $\approx 7.0 \times 10^{10}$ cm
- Earth-Sun distance $\equiv$ 1AU $\approx 1.5 \times 10^{13}$ cm

To go beyond this point, it is useful to adopt a new unit of distance called parsec

$$1 \text{ pc} = \frac{1 \text{ AU}}{1 \text{ arcsec}} = 3.26 \text{ lightyear} = 3 \times 10^{18} \text{ cm} \tag{1.1}$$

The unit of parsec originates in the context of parallax distance-measurements of stars. As shown in the Figure 1.1, if we know the arc length s and the angle θ, the distance r to the star can be inferred. The distance to a star 1 pc is if the length of the arc is 1 AU and the angle it subtends is 1 arcsec.

- A galaxy can be thought of as a disk of radius ~ 30 kpc and thickness ~ 1 kpc (see Figure 1.1)
- The galaxies in our universe are clumped together in clusters of typical size ~ 10 Mpc. The Virgo galaxy cluster, for example, consists of $\sim 10^3 - 10^4$ galaxies.
- Finally, the size of our observable universe is $\sim 5 \times 10^3$ Mpc.

Two of the most fundamental assumptions in cosmology are that the universe is isotropic and homogeneous on large scales, i.e. much larger than the galaxy cluster scale. The isotropy of the universe can be verified by counting and comparing the number of galaxies in different directions or by observing the cosmic microwave background (CMB). The homogeneity of the universe,

Figure 1.1: Measuring the distance to a star with the method of parallax.

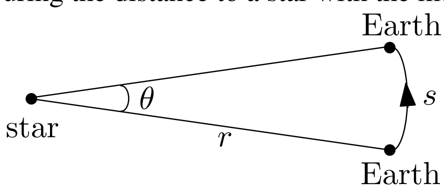
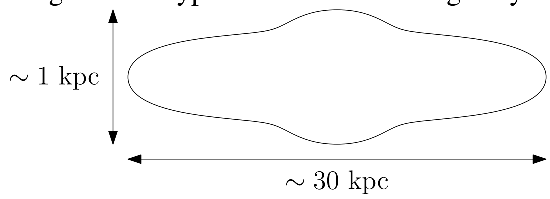


Figure 1.2: Typical dimensions of a galaxy.



on the other hand, is much more difficult to justify as it is practically impossible for us to make measurements at places located at over hundreds of megaparsecs away from us. Nevertheless, we can instead measure the distances between galaxies, reconstruct a 3d picture of the universe out of them, and see if it is homogeneous.

Armed with these assumptions, we now attempt to go as far as we can in constructing a mathematical description of the universe.

1.2 FRW metric

Our first task is to derive the most general homogeneous and isotropic metric. For the sake of clarity, let us forget about the time in spacetime for the moment; we will introduce it back later. A line element in a 3d space is written in terms of the spatial coordinates x^i as

$$d\ell^2 = \gamma_{ij} dx^i dx^j \tag{1.2}$$

with $i, j = 1, 2, 3$ and the signature of the metric γ_{ij} taken to be (+++). The metric γ_{ij} completely determines the geometry of such a space, but it is weakly constrained. A good starting point is the more constrained Riemann curvature tensor $R_{ijk\ell}$.

Reminder:

The Riemann tensor $R_{ijk\ell}$ has the following (anti)symmetry properties:

$R_{ijk\ell} = -R_{jik\ell}$,	antisymmetric in the first two indices
$R_{ijk\ell} = -R_{ij\ell k}$,	antisymmetric in the last two indices
$R_{ijk\ell} = R_{k\ell ij}$,	symmetric in the first-two and last-two pairs of indices

At a particular point, we have sufficient freedom to choose a coordinate system in which

$$\gamma_{ij} = \delta_{ij}, \qquad \partial_k \gamma_{ij} = 0 \tag{1.3}$$

where, as usual, δ_{ij} is the Euclidean metric. The only way to construct a Riemann tensor without breaking isotropy is by building it solely out of δ_{ij}'s (recall that δ_{ij}'s are invariant under rotation)

$$R_{ijk\ell} = \zeta(x) \left(\delta_{ik}\delta_{j\ell} - \delta_{i\ell}\delta_{jk} \right) \tag{1.4}$$

At this point we have solved half of the problem. The homogeneity requirement can be fulfilled by requiring that the proportionality factor $\zeta(x)$ is the same everywhere, i.e. $\zeta(x) = \text{constant}$. Substituting back $\gamma_{ij} = \delta_{ij}$, we arrive at a tensor equation

$$R_{ijk\ell} = \zeta \left(\gamma_{ik}\gamma_{j\ell} - \gamma_{i\ell}\gamma_{jk} \right) \tag{1.5}$$

which is valid in any coordinate system. The above Riemann tensor can be compared with its definition in terms of the metric and its first two derivatives, giving us a 2nd order differential equation which can in principle be solved for γ_{ij}. However, instead of doing that we will follow an easier path.

It follows from (1.5) that

$$R_{ij} = 2\zeta\gamma_{ij}, \qquad \text{Ricci tensor} \tag{1.6}$$
$$R = 6\zeta, \qquad \text{scalar curvature} \tag{1.7}$$

Depending on the value of ζ, the space is qualitatively different:

$$\zeta > 0: \qquad \text{constant positive curvature} \tag{1.8}$$
$$\zeta = 0: \qquad \text{flat} \tag{1.9}$$
$$\zeta < 0: \qquad \text{constant negative curvature} \tag{1.10}$$

It is well known that in the $\zeta = 0$ case the metric can always be chosen to be $\gamma_{ij} = \delta_{ij}$ everywhere (simplest choice). To obtain the metric for $\zeta \neq 0$ cases we can borrow the same idea. We do so by first embedding the space of interest in a flat space of one more dimension, in which the metric can be written down trivially. The induced metric is then obtained by removing one of its dimensions. For example, take a 3-dimensional sphere S^3 and embed it in a 4-dimensional flat space. In this space, the 3-sphere is defined by the equation

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = a^2 \tag{1.11}$$

which is obviously homogeneous and isotropic. Its metric is given by

$$d\ell^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 \tag{1.12}$$

To obtain the induced metric, we seek to eliminate the x_4 dependence in the above metric. Taking the differential of the constraint (1.11)

$$x_1 dx_1 + x_2 dx_2 + x_3 dx_3 + x_4 dx_4 = 0 \tag{1.13}$$

and using it to eliminate the dx_4 in (1.12), we get

$$d\ell^2 = dx_1^2 + dx_2^2 + dx_3^2 + \frac{(x_1 dx_1 + x_2 dx_2 + x_3 dx_3)^2}{a^2 - x_1^2 - x_2^2 - x_3^2} \tag{1.14}$$

The metric (1.14) is simpler when expressed in terms of the spherical coordinates

$$x^1 = r\sin\theta\cos\phi \tag{1.15}$$
$$x^2 = r\sin\theta\sin\phi \tag{1.16}$$
$$x^3 = r\cos\theta \tag{1.17}$$

which allow us to rewrite

$$dx_1^2 + dx_2^2 + dx_3^2 = dr^2 + r^2 d\Omega^2 \tag{1.18}$$
$$x^1 dx_1 + x^2 dx_2 + x^3 dx_3 = r dr \tag{1.19}$$

where

$$d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2 \tag{1.20}$$

All in all, the metric (1.14) can be rewritten as

$$\begin{aligned} d\ell^2 &= dr^2 + r^2 d\Omega^2 + \frac{r^2 dr^2}{a^2 - r^2} \\ &= \frac{dr^2}{1 - \frac{r^2}{a^2}} + r^2 d\Omega^2 \\ &= a^2 \left(\frac{d\bar{r}^2}{1 - \bar{r}^2} + \bar{r}^2 d\Omega^2 \right) \end{aligned} \tag{1.21}$$

In the last step, we have introduced a new dimensionless variable

$$\bar{r} = \frac{r}{a} \tag{1.22}$$

As can be seen from the metric (1.21), in this example of a 3-sphere, whose scalar curvature is positive, the domains of the coordinates are

$$\bar{r} \in [0, 1) \tag{1.23}$$
$$\theta \in [0, \pi] \tag{1.24}$$
$$\phi \in (0, 2\pi) \tag{1.25}$$

Since the domain of $\bar{r}$ is bounded, a positive curvature space is sometimes called *closed*.

By dimensional analysis, we know that the Ricci scalar is proportional to a^{-2}. Therefore, the metric for the negative curvature case can be obtained from the positive curvature one (1.21) by the following substitution

$$a^2 \to -a^2 \tag{1.26}$$

which brings us to

$$\begin{aligned} d\ell^2 &= \frac{dr^2}{1 + \frac{r^2}{a^2}} + r^2 d\Omega^2 \\ &= a^2 \left(\frac{d\bar{r}^2}{1 + \bar{r}^2} + \bar{r}^2 d\Omega^2 \right) \end{aligned} \tag{1.27}$$

As we can see from the above metric, the domains of the coordinates in this case are

$$\bar{r} \in [0, \infty) \tag{1.28}$$
$$\theta \in [0, \pi] \tag{1.29}$$
$$\phi \in (0, 2\pi) \tag{1.30}$$

Since the domain of $\bar{r}$ is unbounded, a negative curvature space is sometimes called *open*.

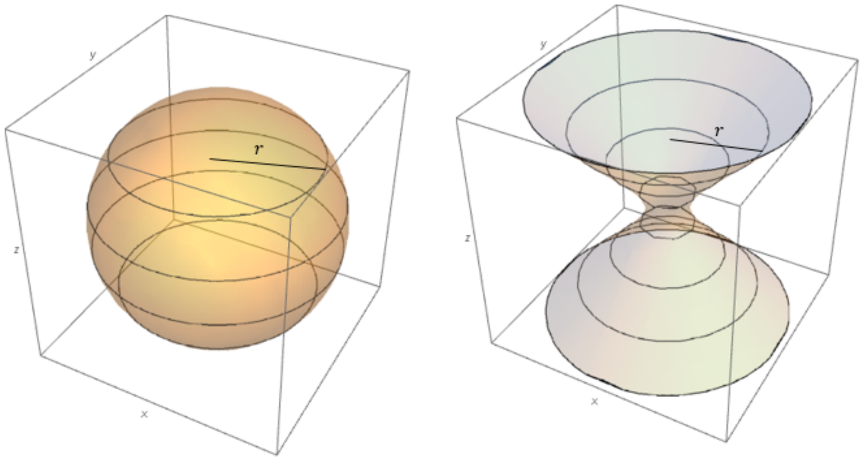
The most general homogeneous and isotropic spatial metric can be written in the following compact form that accounts for all cases (positive, zero, and negative curvature):

$$d\ell^2 = a^2 \left(\frac{d\bar{r}^2}{1 - K\bar{r}^2} + \bar{r}^2 d\Omega^2 \right) \tag{1.31}$$

where

$$K = \begin{cases} +1, & \text{closed, positive curvature} \\ 0, & \text{flat, zero curvature} \\ -1, & \text{open, negative curvature} \end{cases} \tag{1.32}$$

Figure 1.3: Three-dimensional representation of 2-sphere (left) and one-sheet 2-hyperboloid (right).



Restoring the time and accounting for the possibility that a can be time-dependent, we obtain

$$ds^2 = dt^2 - a^2(t)\left(\frac{d\bar{r}^2}{1-K\bar{r}^2} + \bar{r}^2 d\Omega^2\right) \tag{1.33}$$

which is known as the Friedmann-Lemaitre-Robertson-Walker metric or simply the FLRW metric or even FRW metric.

Remarks:

- The FRW metric (1.33) describes all possible homogeneous and isotropic spacetime locally, but not necessarily globally. This is because the topology of the spacetime of interest is not necessarily simple. For example, a spacetime that is locally flat may look like a torus globally. In some cases, it might be possible to identify the topology of a spacetime from within. If, for instance, the galaxies located at different distances away appear to be repeated, then it could be an evidence that the topology of the universe is toroidal.
- It is not always possible to embed a curved space in a flat space of higher dimension. For example, it is known that we cannot embed a negative curvature 2d space in a 3d flat space. We can instead embed it in a 5d or 6d flat space. For a 4d flat space, it is known that we can embed a negative curvature 2d space in it locally, but we still do not know if it is doable globally. Furthermore, it has been proven that a hyperbolic k-space can be embedded in a 5k-5 dimensional flat space or any flat space of dimension $\geq 2k-1$, e.g. a hyperbolic 3-space can be embedded in a 10 dimensional flat space or any flat space of dimension ≥ 7.
- By looking at their metrics, we found that a positively-curved homogeneous and isotropic space is finite (compact) while a negatively-curved one is infinite (non-compact). One way to gain an intuition on this is by studying curved spaces that we can imagine. 2-sphere and (one-sheet) 2-hyperboloid suggest themselves as representatives of $K = +1$ and $K = -1$ spaces respectively. Figure. 1.2 shows how they look like when embedded in a 3d Euclidean space. Following the procedure for obtaining the induced metric we learned in this lecture, we remove the z coordinate. From the pictures, it is clear that r, which corresponds to the projection of the radius onto the xy plane, is bounded for 2-sphere and unbounded for 2-hyperboloid.

Summary:
The size of the observable universe is $\sim 5 \times 10^3$ Mpc. It appears to be homogeneous and isotropic on scales larger than that of the galaxy clusters ($\gtrsim 10$ Mpc). The most general metric compatible with homogeneity and isotropy is given by the FRW metric (1.33)

Lecture 2

Friedmann equations; Friedmann equations in Newtonian gravity; Einstein's static universe; expanding universe.

2.1 Friedmann equations

Having found the FRW metric (1.33) describing a time-dependent homogeneous and isotropic universe, we now turn to its dynamical aspects. We would like to write down the Einstein equation describing such a universe

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{2.1}$$

where $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$ is the Einstein tensor, Λ is the cosmological constant, and $T_{\mu\nu}$ is the energy-momentum tensor. The nonzero components of the Christoffel symbol in Cartesian coordinates are

$$\Gamma^0_{ij} = -\frac{\dot{a}}{a} g_{ij} \tag{2.2}$$

$$\Gamma^i_{0j} = \frac{\dot{a}}{a} \delta^i_j \tag{2.3}$$

$$\Gamma^i_{jk} = \frac{K}{a^2} g_{jk} x^i \tag{2.4}$$

from which we can construct the Ricci tensors and scalar

$$R_{00} = -3\frac{\ddot{a}}{a} \tag{2.5}$$

$$R_{ij} = -\left(\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + \frac{2K}{a^2}\right) g_{ij} \tag{2.6}$$

$$R = -6\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{K}{a^2}\right) \tag{2.7}$$

As a first approximation, we can assume that the universe is a perfect fluid. It's energy-momentum tensor is then

$$T_{\mu\nu} = (\rho + p) u_\mu u_\nu - p g_{\mu\nu} \tag{2.8}$$

with ρ, p, and u_μ being its the energy density, pressure, and 4-velocity of the fluid respectively. The universe is thus not boost invariant and has a preferred cosmic rest frame, namely the rest frame of the fluid, in which $u_\mu = (1,0,0,0)$. In this coordinate system, the non-zero components of the stress energy tensor are

$$T_{00} = \rho \tag{2.9}$$

$$T_{ij} = -p g_{ij} \tag{2.10}$$

Plugging (2.5), (2.7), and (2.9) into the 00^{th} component of the Einstein equation, we get

$$\frac{\dot{a}^2}{a^2} + \frac{K}{a^2} - \frac{\Lambda}{3} = \frac{8\pi G}{3}\rho \tag{2.11}$$

and plugging (2.6), (2.7), and (2.10), into the ij^{th} components of the Einstein equation, we get

$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{K}{a^2} - \Lambda = -8\pi G p \tag{2.12}$$

All the other components of the Einstein equation are identically zero. Note that as a consequence of the homogeneity and isotropy of the universe, the ten Einstein equations have been reduced to only two. Equations (2.11) and (2.12) are important in cosmology. They are known as the *Friedmann equations*. The two Friedmann equations have three unknowns: ρ, p, and a, i.e. they are under determined. In order for them to be solvable, one more equation is needed. The extra equation complementing them is usually taken to be the equation of state of the universe: an equation of the form $\rho = wp$, where w depends on the properties of the energy component in question. Below are three types of energy components that are of special importance in Cosmology together with their equations of states:

- Non-relativistic matter / dust: $p = 0$
- Relativistic matter / radiation: $p = \rho/3$
- Cosmological constant / dark energy: $p = -\rho$

If several species are present, the equation state is given by the sum of each component

$$p = \sum_i w_i \rho_i \tag{2.13}$$

The Friedmann equations also come in several other forms, which are useful in different contexts. For instance, we can take the difference between (2.11) and (2.12) to get a useful form of the Friedmann's equations that does not involve K

$$\frac{\ddot{a}}{a} = \frac{\Lambda}{3} - \frac{4\pi G}{3}(\rho + 3p) \tag{2.14}$$

This form is also useful because it readily tells us about whether the expansion of the universe is accelerating or decelerating. To see other variants of the Friedmann equations, let us consider the energy-momentum conservation

$$\nabla_\nu T^{\mu\nu} = \partial_\nu T^{\mu\nu} + \Gamma^\mu_{\nu\beta} T^{\nu\beta} + \Gamma^\nu_{\nu\beta} T^{\mu\beta} = 0 \tag{2.15}$$

and set $\mu = 0$ to arrive at

$$\partial_\nu T^{0\nu} + \Gamma^0_{\nu\beta} T^{\nu\beta} + \Gamma^\nu_{\nu\beta} T^{0\beta} = 0 \tag{2.16}$$

Feeding it with (2.2), (2.3), (2.9), and (2.10), we get

$$\dot{\rho}+3\frac{\dot{a}}{a}\rho+3\frac{\dot{a}}{a}p=0 \tag{2.17}$$

which could be rewritten in a more physically-transparent form as

$$\frac{\partial}{\partial t}\left(\rho a^3\right)+p\frac{\partial a^3}{\partial t}=0 \tag{2.18}$$

This is just the 1st law of thermodynamics

$$dE+pdV=0 \tag{2.19}$$

with $V \propto a^3$ and $E \propto \rho a^3$. Since the conservation of energy is a consequence of the Bianchi identities, we could have arrived at (2.18) from the Einstein equations via the Bianchi identities. Due to this very reason, we stress here that the energy conservation equation (2.18) is not independent from the two Friedmann equations that we have obtained earlier, namely (2.11) and (2.12), as they are nothing but the components of the Einstein equation. Thus, (2.18) can be regarded as yet another alternate form of one of the Friedmann equations. Having four equations (2.11), (2.12), (2.14), and (2.18) in our hands, we can pick any two equations to form a pair of Friedmann equations which can be solved with the help of an equation of state. For example, if we want to have a set of differential equations that do not involve second derivatives ($\ddot{a}$), then we can pick (2.11) and (2.18) as our choice of Friedmann equations.

2.2 Friedmann equations in Newtonian gravity

General relativity is not necessary in describing a flat space ($K=0$) containing only non-relativistic matter ($p=0$) in the absence of a cosmological constant ($\Lambda=0$). Hence, we expect the Friedmann equations in this case to be derivable from Newtonian gravity. Let us check if this is indeed true. Consider a uniform distribution of matter with energy/mass density ρ. The total energy/mass inside a spherical region of radius a must conserved, so

$$\begin{aligned}\frac{d}{dt}\left(\rho\frac{4}{3}\pi a^3\right)&=0\\ \frac{d}{dt}\left(\rho a^3\right)&=0\end{aligned} \tag{2.20}$$

which coincides with the energy conservation equation (2.18) for $p=0$, as expected. Now, let us study the motion of a test mass m located on the surface of the sphere. Newton's laws of motion and gravity give

$$\begin{aligned}m\ddot{a}&=-G\frac{1}{a^2}\left(\rho\frac{4}{3}\pi a^3\right)m\\ \frac{\ddot{a}}{a}&=-\frac{4}{3}\pi G\rho\end{aligned} \tag{2.21}$$

where we have used the fact that the gravitational field outside of a sphere is equivalent to that due to a point mass located at the center of the sphere and vanishes inside of it. Again, we find that this matches with the special case of (2.14) where $K=0$, $p=0$ and $\Lambda=0$.

2.3 Einstein's static universe

Einstein (and most people during his time) thought that the universe must be static in the sense that $\dot{a} = 0$. Starting from this assumption, he then postulated the existence of cosmological constant. Let us retrace the (hand-wavy) line of reasoning that led him to this conclusion. At that time, the cosmological constant had not been discovered, so let us begin by setting $\Lambda = 0$. The static ($\dot{a} = 0$) Friedmann equations (2.11) and (2.12) then reduce to

$$\frac{K}{a^2} = \frac{8\pi G}{3}\rho \tag{2.22}$$
$$\frac{K}{a^2} = -8\pi G p \tag{2.23}$$

Einstein observed that the universe is not empty (there are stars!), so its energy density must be non-zero

$$\rho \neq 0 \tag{2.24}$$

Furthermore, he observed that stars do not move, so

$$p = 0 \tag{2.25}$$

If these observations are true then (2.22) and (2.23) would be inconsistent; the former gives $K \neq 0$ while the latter gives $K = 0$. To reconcile these, in 1917 Einstein realized that the presence of a cosmological constant Λ in the Einstein equation would appear in the Friedmann equations as

$$\frac{K}{a^2} - \frac{\Lambda}{3} = \frac{8\pi G}{3}\rho \tag{2.26}$$
$$\frac{K}{a^2} - \Lambda = 0 \tag{2.27}$$

and would allow for a consistent solution given by

$$a = \frac{1}{\sqrt{\Lambda}}, \qquad K = +1, \qquad \Lambda = 4\pi G\rho \tag{2.28}$$

The presence of such a cosmological constant gives rise to an apparent conceptual problem. If we set $\rho = 0$ and $\Lambda \neq 0$, then the Friedmann equations (2.11) and (2.12) only accept non-trivial solutions where the space is curved $K \neq 0$ and non-static $\dot{a} \neq 0$. In other words, even an empty space has a non-trivial dynamics (and is curved). This is actually not a real physical problem. It is just a possibility that people tend to find difficult to accept. After Friedmann published his model, which we will discuss in the next section, in 1922, Einstein stated that introducing the cosmological constant Λ was his biggest blunder.

2.4 Expanding universe

Unlike Einstein, Friedmann thought that it is more reasonable that the universe has no cosmological constant $\Lambda = 0$, but is dynamical. Here, for no other reason than simplicity, let us set $K = 0$ and $p = 0$, i.e. a matter dominated universe, so that the Friedmann's equations (2.11) and (2.12) reduce to

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho \tag{2.29}$$
$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} = 0 \tag{2.30}$$

Integrating the second equation, we get

$$\begin{aligned} 2\int dt\frac{\ddot{a}}{\dot{a}}+\int dt\frac{\dot{a}}{a}=0 \\ 2\log\dot{a}+\log a=C \\ a^2\dot{a}=C \end{aligned} \tag{2.31}$$

where C is an integration constant. Integrating once more yields

$$\begin{aligned} \int a^{1/2}da=\int Cdt \\ a^{3/2}=Ct \\ a=a_0\left(\frac{t}{t_0}\right)^{2/3} \end{aligned} \tag{2.32}$$

where we have neglected the new integration constant, as it is negligible for large enough t's. Using the obtained $a(t)$, we can calculate how the energy density evolves with time with the help of (2.29)

$$\rho=\frac{3}{8\pi G}\frac{\dot{a}^2}{a^2}=\frac{1}{6\pi G}\frac{1}{t^2} \tag{2.33}$$

A prediction of Friedmann's model is that the distances between stars must increase with time. This can be seen in the following way. Consider two objects located at fixed coordinate points (such objects are called comoving) separated by a constant coordinate distance $\Delta\bar{\ell}$ away. The physical distance $\Delta\ell$ between them is given by

$$\begin{aligned} \Delta\ell^2=a^2(t)\Delta\bar{\ell}^2 \\ \Delta\ell=a(t)\Delta\bar{\ell} \end{aligned} \tag{2.34}$$

Taking the time-derivative while keeping the coordinate distance $\Delta\bar{\ell}$ fixed

$$\frac{d\Delta\ell}{dt}=\dot{a}(t)\Delta\bar{\ell} \tag{2.35}$$

and rewriting $\Delta\bar{\ell}$ in terms of $\Delta\ell$ using (2.34), we get

$$\frac{d\Delta\ell}{dt}=\frac{\dot{a}(t)}{a(t)}\Delta\ell \tag{2.36}$$

By defining the *Hubble parameter / Hubble constant*

$$H=\frac{\dot{a}}{a} \tag{2.37}$$

we can rewrite (2.36) as

$$\frac{d\Delta\ell}{dt}=H\Delta\ell \tag{2.38}$$

which is known as the *Hubble's law*. It says that two objects located at a constant coordinate separation $\Delta\bar{\ell}$ away from each other move apart with a physical speed $d\Delta\ell/dt$ proportional to their physical distance $\Delta\ell$. For instance, if we plug in the solution for a matter dominated universe (2.32), we obtain

$$\frac{d\Delta\ell}{dt} = \frac{2}{3t}\Delta\ell \tag{2.39}$$

To verify the Hubble's law experimentally, we need a way to measure the distances of faraway objects. One way to do so is by comparing the inherent luminosities of the so called *standard candles* (stars with known luminosities) with the apparent ones. By observing how the distances of these standard candles evolve with time, we can check the validity of the Hubble's law. The receding speeds of stars can be measured more accurately by measuring the shifts in their spectral lines due to the Doppler effect. According to the Doppler effect, the frequency of a photon emitted with frequency ω_0 by a star that is moving at a speed v away from us is redshifted to

$$\omega = \sqrt{\frac{1-v}{1+v}}\omega_0 \tag{2.40}$$

By measuring the redshifts in the spectra of various stars, Hubble managed to obtain a distance vs velocity relation that shows a general trend of velocity increasing with distance. Despite the highly scattered data points he obtained, Hubble was bold enough to draw a straight line corresponding to $H = 500\frac{\text{km/s}}{\text{Mpc}}$ in his 1929 paper. This differs by about an order of magnitude from the results of more precise recent measurements, settling at around $H = 70\frac{\text{km/s}}{\text{Mpc}}$. Even to this day, the best measurements of the Hubble parameter still contains high degrees of error, mainly stemming from our inability of measuring the distances of faraway objects precisely.

Summary:
Friedmann equations are two equations that, as a pair, are equivalent to the Einstein equations for an FRW universe filled with perfect-fluid. Friedmann equations come in many forms. Four of them (2.11), (2.12), (2.14), and (2.18) are derived here. Einstein found that if we assume that the universe is static, has nonzero energy density, and has zero pressure, then the Friedmann equations would be inconsistent unless the cosmological constant is nonzero. Friedmann, having come to peace with the possibility that the universe is dynamical, found that (as an example) a matter dominated universe would expand with time. In such a universe, two objects located at fixed coordinate positions move apart according to the Hubble's law, i.e. with a physical speed that is proportional to their physical distance.

Lecture 3

> redshift in an FRW universe; definition of luminosity distance; luminosity distance and Hubble's law

3.1 Redshift in an FRW universe

Suppose that a light pulse is emitted at $t = t_1$ by a star located at the comoving coordinates $\theta = \pi/2$, $\phi = 0$, and $\bar{r} = \bar{r}_1$ in an FRW universe (described by the metric (1.33)) to be detected by us at the same angular coordinates at $t = t_0$ and $\bar{r} = 0$. On its way to our detector, the pulse follows the null geodesic

$$0 = ds^2 = dt^2 - a^2(t)\frac{d\bar{r}^2}{1 - K\bar{r}^2} \tag{3.1}$$

Taking the square root and integrating gives

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_0^{\bar{r}_1} \frac{d\bar{r}}{\sqrt{1 - K\bar{r}^2}} \equiv f(\bar{r}_1) = \begin{cases} \arcsin \bar{r}_1, & K = +1 \\ \bar{r}_1, & K = 0 \\ \text{arcsinh}\ \bar{r}_1, & K = -1 \end{cases} \tag{3.2}$$

where we have picked the negative root since we are interested in light pulse that travels in the decreasing instead of increasing $\bar{r}$ direction. We will find this equation useful in future discussions.

Suppose that a second pulse is emitted by the star shortly after the first one at $t = t_1 + \delta t_1$ and detected at $t = t_0 + \delta t_0$ by our detector. As the two pulses are both emitted at the same comoving radial coordinate $\bar{r}_1$, (3.2) yields

$$\begin{aligned} f(\bar{r}_1) &= f(\bar{r}_1) \\ \int_{t_1+\delta t_1}^{t_0+\delta t_0} \frac{dt}{a(t)} &= \int_{t_1}^{t_0} \frac{dt}{a(t)} \end{aligned} \tag{3.3}$$

For sufficiently small δt_0 and δt_1, we have

$$\frac{\delta t_0}{a(t_0)} = \frac{\delta t_1}{a(t_1)} \tag{3.4}$$

If the two pulses coincide with the peaks of an electromagnetic wave, we can substitute $\delta t_0/\delta t_1 = \omega_1/\omega_0$, where ω_0 and ω_1 are the observed and emitted frequency respectively, to get

$$\omega_0 = \omega_1 \frac{a(t_1)}{a(t_0)} \tag{3.5}$$

or, in terms of wavelengths,

$$\lambda_0 = \lambda_1 \frac{a(t_0)}{a(t_1)} \tag{3.6}$$

Hence, the wavelength of a photon rescales just like how the physical distance between any two points would. If the universe is expanding, we have $a(t_0) > a(t_1)$, and so $\lambda_0 > \lambda_1$ (redshift). On the other than, if the universe is shrinking, we have $a(t_0) < a(t_1)$, and so $\lambda_0 < \lambda_1$ (blueshift). In astrophysics, the amount of redshift / blueshift is often measured in terms of a quantity

$$z = \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{a(t_0)}{a(t_1)} - 1 \tag{3.7}$$

which we simply call the *redshift*. It is often useful to express results in terms of the redshift z as it is one of a few quantities that we can directly measure.

3.2 Luminosity distance and Hubble's law

The Hubble's law (2.38) we found in the previous lecture states that in an FRW universe two comoving points move apart with a physical speed that is proportional to their physical distance. We immediately run into a problem when we decide to confirm this statement experimentally. The reason is that at cosmological scales the physical distance between two points at an instance is practically impossible to measure. We therefore need to resort to indirect methods of measuring distance, which typically involve detecting light from faraway objects. The price to pay is that the said indirect methods do not actually measure the physical distance at a particular moment in time, but more complicated notions of distance that depend on the properties of the universe at different moments in time during the light's journey. Examples of these indirect methods include:

- Comparing the apparent luminosities of *standard candles* (objects with known luminosities) to their intrinsic luminosities.
- Comparing the apparent sizes of *standard rulers* (objects with known sizes) to their actual sizes.
- The method of parallax: measuring the shifts in the apparent positions of the objects of interest due to the orbital motion of the Earth.

Let us study the first method in more detail, as it is the method that works most effectively for the largest distances. This method leads to a notion of distance called the *photometric distance* or *luminosity distance*. The idea is that, in the absence of expansion, the apparent luminosity (energy flux per unit time) P registered at a telescope pointing towards a spherically symmetric light source whose intrinsic luminosity is L is given by

$$P = L\frac{S}{4\pi d^2} \tag{3.8}$$

where S is the cross-section of the telescope and d in this case is the distance between the light source and the telescope. Inspired by this, we define the luminosity distance as

$$d = \sqrt{\frac{LS}{4\pi P}} \tag{3.9}$$

though in an expanding universe d would no longer be the same as the source-telescope distance. Since S and L are typically known, we can measure d by measuring P.

Now, we shall try to understand what information we actually get when we measure the luminosity distance d. To this end, we derive the apparent luminosity P from the intrinsic luminosity L. P can be obtained from L by multiplying it with three factors. First, there is the geometrical fraction factor that is due to the fact that only a fraction of the light emitted by the light source is captured by the telescope. At the time of detection t_0, the light front emitted at time t_1 has spread to a sphere of comoving radius $\bar{r}_1$ which can obtained by inverting (3.2) to solve for $\bar{r}_1$. To calculate the area of this light front, consider the space interval at $t = t_0$

$$d\ell^2 = a^2(t_0)\left[\frac{d\bar{r}^2}{1-K\bar{r}^2} + \bar{r}^2 d\Omega^2\right] \tag{3.10}$$

The light front corresponds to the surface of constant $\bar{r}$. If we set $d\bar{r} = 0$, the above spatial metric is equivalent to that of a flat space. So, the total area of the wavefront is simply given by

$$S_{\text{tot}} = \int a^2(t_0)\bar{r}_1^2 d\Omega^2 = 4\pi\bar{r}_1^2 a^2(t_0) \tag{3.11}$$

Out of this light front, only a small fraction S/S_{tot}, with S being the cross-section of the telescope, is captured by the telescope. Second, as we discussed in the previous section, the observed frequency of the light emitted by the light source is redshifted compared to the original frequency, and as a result the luminosity is modified by the ratio of the observed and emitted photon energies $(\hbar\omega_0)/(\hbar\omega_1)$. Third, due to the difference in the metric at the point of emission and detection, the time-flow rate differ at the two points. Consequently, the luminosity is further modified by the ratio of the two rates $\delta t_1/\delta t_0$. Putting all the factors together, we find

$$P = L\left(\frac{S}{S_{\text{tot}}}\right)\left(\frac{\hbar\omega_0}{\hbar\omega_1}\right)\left(\frac{\delta t_1}{\delta t_0}\right) \tag{3.12}$$

Noting that S_{tot} is given by (3.11), $\omega_0/\omega_1 = a_1/a_0$, and $\delta t_0/\delta t_1 = a_0/a_1$ (as found in the previous section), the observed luminosity is given by

$$P = L\frac{a^2(t_1)}{a^2(t_0)}\frac{S}{4\pi a^2(t_0)\bar{r}_1^2} \tag{3.13}$$

Comparing this with (3.8), we find the information content packaged in the luminosity distance

$$d = a^2(t_0)\frac{\bar{r}_1(t_0,t_1)}{a(t_1)} \tag{3.14}$$

This can be taken as the theoretical definition of the luminosity distance.

With the help of (3.2) and (3.7) we can relate d to z. As a warm up here we consider the case where $\bar{r}_1$ is small and so t_1 is close to t_0; we will do it generally in the next lecture. In this case, (3.2) reduces to

$$\frac{t_0 - t_1}{a(t_0)} \approx \bar{r}_1 \tag{3.15}$$

since $a(t)$ is approximately constant and $f(\bar{r}_1) \approx \bar{r}_1$. Furthermore, by Taylor expanding $a(t_1) \approx$

$a(t_0) + \dot{a}_0(t_1 - t_0)$ in (3.7), we get

$$
\begin{aligned}
z &\approx \frac{a(t_0)}{a(t_0) + \dot{a}_0(t_1 - t_0)} - 1 \\
&\approx \frac{\dot{a}(t_0)(t_0 - t_1)}{a(t_0)} \\
&\approx \frac{\dot{a}(t_0)}{a(t_0)} \bar{r}_1 a(t_0)
\end{aligned} \tag{3.16}
$$

where in the last step we employed (3.15). Now, we can use (3.16) to express the $\bar{r}_1$ in (3.14), which is now reduced to $d \approx a(t_0)\bar{r}_1$, in terms of z

$$
z \approx \frac{\dot{a}(t_0)}{a(t_0)} d = Hd \tag{3.17}
$$

This is the Hubble's law written in terms of experimentally measurable quantities.

Summary:
In an expanding universe, the observed frequency of light emitted by a faraway object is redshifted relative to its emitted frequency. The amount of redshift is often measured in terms of a quantity z, defined in (3.7). At cosmological scales, the physical distance is practically difficult to measure. The luminosity distance is an alternative notion of distance that is measurable. It is defined in terms of experimentally measurable quantities in (3.9) and in terms of unmeasurable, but theoretically useful quantities in (3.14). As found in (3.17), the Hubble's law can be written in terms measurable quantities.

Lecture 4

Relation between luminosity distance and redshift; critical density; content of the universe

4.1 Relation between luminosity distance and redshift (the plan)

In the previous lecture, we defined two experimentally measurable quantities, the luminosity distance d and redshift z. Our goal in this lecture is to relate the two quantities. The reason we are interested in the relation between d and z is because it depends on the content of the universe. This means that by measuring it we can probe the content of the universe. From the definitions of luminosity distance (3.14) and redshift (3.7), we have

$$d = \bar{r}_1(t_1, t_0) a(t_0)(1+z) \tag{4.1}$$

We want to write the everything on the right hand side solely in terms of z. First, we use (3.2) to express $\bar{r}_1(t_1, t_0)$ in terms of t_1

$$d = \begin{cases} a(t_0)(1+z)\sin\left[\int_{t_1}^{t_0} \frac{dt}{a(t)}\right], & K = +1 \\ a(t_0)(1+z)\int_{t_1}^{t_0} \frac{dt}{a(t)}, & K = 0 \\ a(t_0)(1+z)\sinh\left[\int_{t_1}^{t_0} \frac{dt}{a(t)}\right], & K = -1 \end{cases} \tag{4.2}$$

What is left to be done is to express the integral

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} \tag{4.3}$$

in terms of z. In order to do that, we need to solve for the time-dependence of the scale factor $a(t)$. In this lecture, we will do exactly that, while deriving useful results in the process.

4.2 Critical density

The term involving the cosmological constant $-\Lambda/3$ in the Friedmann equation (2.11), when moved to the right hand side of the equation, can be interpreted as due to a constant energy density

$$\rho_\Lambda \equiv \frac{\Lambda}{8\pi G} \tag{4.4}$$

which we refer to as the *vacuum energy density*. While the name is based on our best guess for its nature, i.e. that the energy density of vacuum is indistinguishable from a cosmological constant, the right interpretation is not yet known. Similarly, if we move the $-\Lambda$ term in the Friedmann equation (2.12) to the right hand side, we can interpret it as due to the *vacuum pressure*

$$p_\Lambda \equiv -\frac{\Lambda}{8\pi G} \tag{4.5}$$

As we can see,

$$p_\Lambda = -\rho_\Lambda \tag{4.6}$$

So, the vacuum energy component has an equation of state parameter $w_\Lambda = -1$. With these definitions, we can define the total energy as

$$\rho_{\text{tot}} \equiv \rho + \rho_\Lambda \tag{4.7}$$

and the total pressure as

$$p_{\text{tot}} \equiv p + p_\Lambda \tag{4.8}$$

Using the definition of the Hubble parameter (2.37), the Friedmann equation (2.11) and the energy conservation equation (2.17) can be written respectively as

$$H^2 + \frac{K}{a^2} = \frac{8\pi G}{3}\rho_{\text{tot}} \tag{4.9}$$

$$\frac{d\rho_{\text{tot}}}{dt} = -3H(\rho_{\text{tot}} + p_{\text{tot}}) \tag{4.10}$$

From (4.9), we see that the present universe is flat ($K = 0$) if the present total energy density $\rho_{\text{tot},0}$ is equal to the *critical density*

$$\rho_c \equiv \frac{3H_0^2}{8\pi G} \tag{4.11}$$

Though the critical density ρ_c could be defined at any time, throughout these lectures it is *always* defined at the present time $t = t_0$. ρ_c can be found by measuring the Hubble parameter H_0. For $H_0 = 100h$ (km/s)/Mpc, the critical density is $\rho_c = 1.88h^2 \times 10^{-29}\text{g/cm}^3$.

The relative size between the total density today $\rho_{\text{tot},0}$ and the critical density ρ_c tells us about the curvature of the universe:

$$\begin{aligned}
\rho_{\text{tot},0} > \rho_c \quad &\longrightarrow \quad K > 0: \quad \text{the universe is closed} \\
\rho_{\text{tot},0} = \rho_c \quad &\longrightarrow \quad K = 0: \quad \text{the universe is flat} \\
\rho_{\text{tot},0} < \rho_c \quad &\longrightarrow \quad K < 0: \quad \text{the universe is open}
\end{aligned}$$

Apart from that, it also tells us about the fate of the universe (whether it expands forever or bounces back). As an example, consider a matter dominated universe with $p = 0$ and $\Lambda = 0$. In this case, the Friedmann equation (2.14) that is independent of K reduces to

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\rho \tag{4.12}$$

Multiplying both sides with $a\dot{a}$ and using the energy conservation $\rho a^3 = \text{constant} = \rho_0 a_0^3$ (from here and on we will write a_0 and $a(t_0)$ interchangeably), we arrive at

$$\dot{a}\ddot{a} = -\frac{4\pi G}{3}\left(\rho_0\frac{a_0^3}{a^3}\right)a\dot{a} \tag{4.13}$$

Then, integrating over t gives

$$\frac{1}{2}\dot{a}^2 - \frac{4\pi G}{3}\rho_0 a_0^3 \frac{1}{a} = \text{const} = -\frac{4\pi G}{3} a_0^3(\rho_0 - \rho_c) \tag{4.14}$$

where the constant of integration was found by evaluating the left hand side at $t = t_0$. This resembles the energy conservation equation for a particle of unit mass moving under the influence of a potential

$$U(a) = -\frac{4\pi G}{3}\rho_0 a_0^3 \frac{1}{a} \tag{4.15}$$

with total energy

$$E = -\frac{4\pi G}{3} a_0^3(\rho_0 - \rho_c) \tag{4.16}$$

By the particle motion analogy, it is clear that the fate of the universe depends on the relative size between ρ_0 and ρ_c. There are two cases. If $\rho_0 \leq \rho_c$, and hence $E \geq 0$, the universe would expand forever with $a(t) \to \infty$ as $t \to \infty$. On the other hand, if $\rho_0 > \rho_c$, and hence $E < 0$, the universe would expand for a finite amount of time, and then collapse.

4.3 Content of the universe

Suppose that the universe is composed of:

- radiation, with $p_\gamma = \rho_\gamma/3$
- non-relativistic matter (dark and baryonic combined), with $p_m = 0$
- cosmological constant, with $p_\Lambda = -\rho_\Lambda$

This is, in fact, the simplest model of the universe that is broadly consistent with the observed features of the universe. We would like to study how such a universe evolves with time. Before going any further, it will prove useful to express the content of the present universe in terms of the dimensionless numbers $\Omega_i = \rho_{i,0}/\rho_c$, called *abundances*. Explicitly,

$$\Omega_\Lambda \equiv \frac{\rho_\Lambda}{\rho_c}; \quad \Omega_m \equiv \frac{\rho_{m,0}}{\rho_c}; \quad \Omega_\gamma \equiv \frac{\rho_{\gamma,0}}{\rho_c} \tag{4.17}$$

We stress here that, like the critical density ρ_c, the abundances Ω_i in these lectures are defined at the present time $t = t_0$. The Friedmann equation (4.9) can be written as

$$H^2 = \frac{8\pi G}{3}\left(\rho_\gamma + \rho_m + \rho_\Lambda + \rho_K\right) \tag{4.18}$$

where we have defined the curvature contribution to the energy density as

$$\rho_K = -\frac{3}{8\pi G}\frac{K}{a^2} \tag{4.19}$$

By defining the abundance associated to the curvature as

$$\Omega_K \equiv \frac{\rho_{K,0}}{\rho_c} = -\frac{K}{a_0^2 H_0^2} \tag{4.20}$$

and with the use of (4.11) and (4.17), we can write the Friedmann equation (4.18) at the present time as a sum rule

$$1 = \Omega_\Lambda + \Omega_K + \Omega_m + \Omega_\gamma \tag{4.21}$$

This form of the Friedmann equation is compact and useful for consistency checking, but it does not tell us about how the universe expands.

To obtain an equation that dictates the expansion of the universe, we need to understand how the different energy components evolve as the universe expands. For simplicity and solubility, we will assume that the different components do not interact with one another. This assumption is valid at the present epoch, but it becomes increasingly inaccurate as we go backward in time, as the temperature was higher and interactions were stronger then. The energy conservation equation (2.18) for matter

$$\frac{d}{dt}\left(\rho_m a^3\right) = 0 \tag{4.22}$$

implies that the matter energy density scales as

$$\rho_m \propto \frac{1}{a^3} \tag{4.23}$$

Next, the energy conservation equation for radiation gives

$$\begin{aligned} \frac{\partial}{\partial t}\left(\rho_\gamma a^3\right) + p_\gamma \frac{\partial a^3}{\partial t} &= 0 \\ \frac{d}{dt}\left(\rho_\gamma a^4\right) &= 0 \end{aligned} \tag{4.24}$$

where we have used the equation of state for radiation $\rho_\gamma = p_\gamma/3$. The last equation reads

$$\rho_\gamma \propto \frac{1}{a^4} \tag{4.25}$$

which can be understood as follows: the number density of photons scales as $1/a^3$ and the energy/frequency of each photon scales as $1/a$, so together they make $1/a^4$. By definition, the curvature energy density (4.19) scales as

$$\rho_K \propto \frac{1}{a^2} \tag{4.26}$$

Finally, also by definition, the cosmological constant energy density is unchanging

$$\rho_\Lambda = \text{constant} \tag{4.27}$$

Knowing the scaling behaviour of each component, the Friedmann equation (4.18) can be developed as follows

$$\begin{aligned} H^2 &= \frac{8\pi G}{3}\left[\rho_\Lambda + \rho_K + \rho_m + \rho_\gamma\right] \\ &= \frac{8\pi G}{3}\rho_c\left[\frac{\rho_\Lambda}{\rho_c} + \frac{\rho_{K,0}}{\rho_c}\left(\frac{a_0}{a}\right)^2 + \frac{\rho_{m,0}}{\rho_c}\left(\frac{a_0}{a}\right)^3 + \frac{\rho_{\gamma,0}}{\rho_c}\left(\frac{a_0}{a}\right)^4\right] \\ &= H_0^2\left[\Omega_\Lambda + \Omega_K\left(\frac{a_0}{a}\right)^2 + \Omega_m\left(\frac{a_0}{a}\right)^3 + \Omega_\gamma\left(\frac{a_0}{a}\right)^4\right] \end{aligned} \tag{4.28}$$

In the last step, (4.11) and (4.17) were used. Using the definition of the Hubble parameter, we can rewrite it in a more useful form as

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2\left[\Omega_\Lambda + \Omega_K\left(\frac{a_0}{a}\right)^2 + \Omega_m\left(\frac{a_0}{a}\right)^3 + \Omega_\gamma\left(\frac{a_0}{a}\right)^4\right] \tag{4.29}$$

Given the abundances Ω_i and the Hubble parameter H_0 at $t = t_0$, this equation tells us pretty much everything about how the universe expands. It can also be rearranged to a form that is more intuitive

$$\frac{1}{2}\dot{a}^2 + U(a) = 0 \tag{4.30}$$

where

$$U(a) = -\frac{H_0^2}{2}\left(\Omega_\Lambda a^2 + \Omega_K a_0^2 + \Omega_m \frac{a_0^3}{a} + \Omega_\gamma \frac{a_0^4}{a^2}\right) \tag{4.31}$$

(4.30) describes the energy conservation of a particle of unit mass moving under a potential $U(a)$.

4.4 Relation between luminosity distance and redshift (the result)

We are now in a position to fulfill the goal we set in the beginning of the lecture: expressing the luminosity distance d solely in terms of the redshift z. We start by defining $x = a/a_0$ and

$$A^2(x) \equiv \Omega_\Lambda + \Omega_K \frac{1}{x^2} + \Omega_m \frac{1}{x^3} + \Omega_\gamma \frac{1}{x^4} \tag{4.32}$$

which encapsulates the information about the content of the universe, so that we can write (4.29) compactly as

$$\begin{aligned}\left(\frac{\dot{x}}{x}\right)^2 &= H_0^2 A^2(x) \\ \frac{dx}{dt} &= H_0 A(x) x\end{aligned} \tag{4.33}$$

This allows us to express the sought-after integral (4.3) as

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{t_1}^{t_0} \frac{dt}{dx}\frac{dx}{a_0 x} = \frac{1}{a_0 H_0}\int_{\frac{1}{1+z}}^{1} \frac{dx}{x^2 A(x)} \tag{4.34}$$

where we have used (4.33) to rewrite dt/dx in terms of x and the fact that $x(t_0) = a_0/a_0 = 1$ and $x(t_1) = a(t_1)/a(t_0) = 1/(1+z)$ by the definition of redshift (3.7). We can then plug (4.34) to (4.2) to obtain $d(z)$. For example, for $K = 0$ we have

$$d(z) = \frac{1+z}{H_0}\int_{\frac{1}{1+z}}^{1} \frac{dx}{x^2 A(x)} \tag{4.35}$$

All the three cases $K = -1, 0, +1$ can be captured by a single equation

$$d(z) = \frac{1+z}{H_0\sqrt{\Omega_K}} \sinh\left[\sqrt{\Omega_K}\int_{\frac{1}{1+z}}^{1} \frac{dx}{x^2 A(x)}\right] \tag{4.36}$$

Its reduction to each case is more readily seen if we use (4.20) to write out Ω_K explicitly

$$d(z) = a_0(1+z)\frac{1}{\sqrt{-K}} \sinh\left[\frac{\sqrt{-K}}{a_0 H_0}\int_{\frac{1}{1+z}}^{1} \frac{dx}{x^2 A(x)}\right] \tag{4.37}$$

The current form coincides with the expression for $K = -1$. When $K = +1$, we have $\sqrt{-K} = i$, and so the sinh would transform into a sin, as it should. The flat case $K = 0$ is obtained by taking the formal limit $\Omega_K \to 0$.

By measuring $d(z)$ and fitting the curve with the expression (4.37), the content of the universe can be inferred. Using (4.21), we can eliminate Ω_K from the fitting-parameter set, leaving us with Ω_γ, Ω_m, and Ω_Λ. $\rho_{\gamma,0}$ and $\rho_{c,0}$ can be obtained directly from the CMB and Hubble parameter measurements, giving Ω_γ. It turns out that $\Omega_\gamma \ll 1$, i.e. can be neglected in practice. The best-fit values turned out to be $\Omega_m \approx 0.3$ and $\Omega_\Lambda \approx 0.7$. Based on this, the so called Lambda-Cold-Dark-Matter model (ΛCDM model), also known as the standard model of cosmology, acquires its name.

Another virtue of (4.33) is that it allows us to calculate the age of the universe t_0 by integrating it from $t = 0$, corresponding to $x = 0$ (since $a = 0$ at that time), to $t = t_0$, corresponding to $x = a_0/a_0 = 1$

$$t_0 = \frac{1}{H_0}\int_0^1 \frac{dx}{xA(x)} \tag{4.38}$$

Summary:
The critical density ρ_c (4.11) is the value of the present energy density that yields a flat universe. The relative size between ρ_c and the current total energy density $\rho_{\text{tot},0}$ determines the curvature of the present universe and whether it will expand forever or shrink into singularity after expanding for a finite amount of time. The luminosity distance - redshift relation $d(z)$, given in (4.37), depends on the energy content of the universe, which is encapsulated by the function $A(x)$. By measuring $d(z)$, we can probe the energy content of the universe.

Lecture 5

> cosmological constant domination; radiation domination; chronology of cosmological epochs

5.1 Cosmological epochs

In the previous lecture, we found that the expansion of a universe that comprises of matter, radiation, and cosmological constant (the ΛCDM model) is dictated by the non-linear equation (4.29), which is difficult to solve in general. However, most of the time there is a single energy component that dominates far above the others, greatly simplifying the dynamics. It is therefore instructive to study how the universe evolves when it is dominated by each of the energy components of the ΛCDM model, namely matter, radiation, and cosmological constant. Matter-dominated universe was already discussed in Lecture 2, with the key results given in (2.32) and (2.33). Here, we will study the other two and give an overview of the different cosmological epochs.

5.1.1 Cosmological constant dominated universe

Consider a cosmological constant dominated universe, with $\rho = 0$, $p = 0$, $K = 0$, but $\Lambda \neq 0$. The motivation for setting $K = 0$ is that we know that Ω_K is negligible today. It would be even more so in the past as it scales more slowly with respect to a than matter and radiation do, and it will never dominate over the cosmological constant in the future as it decays rather than grows. The Friedmann equations (2.11) and (2.12) in this case are

$$\frac{\dot{a}^2}{a^2} = \frac{\Lambda}{3} \qquad (\text{valid only if } \Lambda > 0) \tag{5.1}$$

$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} - \Lambda = 0 \tag{5.2}$$

The first equation can be written as

$$\frac{\dot{a}}{a} = \sqrt{\frac{\Lambda}{3}} \tag{5.3}$$

The solution is simply

$$a = a_0 \exp\left(\sqrt{\frac{\Lambda}{3}}t\right) \tag{5.4}$$

Taking the second derivative gives

$$\ddot{a} = a_0 \frac{\Lambda}{3} \exp\left(\sqrt{\frac{\Lambda}{3}} t\right) > 0 \tag{5.5}$$

i.e. an accelerated expansion, in contrast to what we have in a matter dominated universe with the solution $a \propto t^{2/3}$, found in (2.32), which gives $\ddot{a} < 0$. Note that since we have set $K = 0$, the above analysis does not account for the possibility that $\Lambda < 0$. We can check that setting $\Lambda < 0$ in (5.1) would yield an imaginary Hubble parameter, which is absurd since the scale factor is a real number.

5.1.2 Radiation dominated universe

A universe dominated by radiation, with $K = 0$ and $\Lambda = 0$, obey the Friedmann equations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho \tag{5.6}$$

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = -8\pi G p \tag{5.7}$$

Substituting the power-law ansatz $a \propto t^\alpha$ to both equations, we get

$$\frac{\alpha^2}{t^2} = \frac{8\pi G}{3}\rho \tag{5.8}$$

$$\frac{2\alpha(\alpha-1)}{t^2} + \frac{\alpha^2}{t^2} = -\frac{8\pi G}{3}\rho \tag{5.9}$$

where we have used the equation of state for radiation $p = \rho/3$ to arrive at the second equation. Eliminating ρ gives an equation for α

$$\alpha(2\alpha - 1) = 0 \tag{5.10}$$

So, either $\alpha = 0$ or $\alpha = 1/2$. Since the former one corresponds to a static universe, only the latter one is of interest. Hence, the solution is given by

$$a = a_0 \left(\frac{t}{t_0}\right)^{1/2} \tag{5.11}$$

and so, according to (5.8), the energy density evolves as

$$\rho = \frac{3}{32\pi G t^2} \tag{5.12}$$

5.1.3 Summary and chronology

Below is a summary of the properties of the universe dominated by each of the energy components of the ΛCDM model

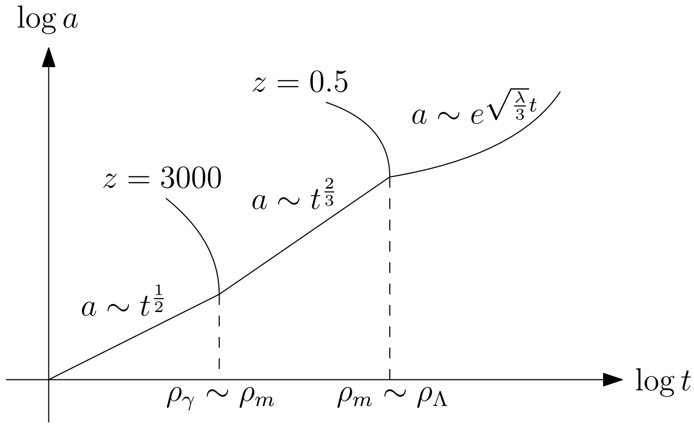
- Matter domination ($p = 0$):

$$a = a_0 \left(\frac{t}{t_0}\right)^{2/3} \tag{5.13}$$

$$\rho = \frac{1}{6\pi G}\frac{1}{t^2} \propto \frac{1}{a^3} \tag{5.14}$$

$$H = \frac{2}{3t} \tag{5.15}$$

Figure 5.1: Chronology of cosmological epochs in the ΛCDM model. As far as ΛCDM model is concerned, the universe was dominated by radiation in its earliest stage. It then went through an intermediary matter-domination stage before finally entering a never-ending cosmological-constant-domination stage.



- Radiation domination ($p = \rho/3$):

$$a = a_0 \left(\frac{t}{t_0}\right)^{1/2} \tag{5.16}$$

$$\rho = \frac{3}{32\pi G}\frac{1}{t^2} \propto \frac{1}{a^4} \tag{5.17}$$

$$H = \frac{1}{2t} \tag{5.18}$$

- Cosmological constant domination ($p = -\rho$):

$$a = a_0 e^{Ht} \tag{5.19}$$

$$\rho = \frac{\Lambda}{8\pi G} = \text{constant} \tag{5.20}$$

$$H = \sqrt{\frac{8\pi G}{3}\rho} = \text{constant} \tag{5.21}$$

Figure 5.1.3 shows rough time-evolution of the scale factor in the ΛCDM model. We are currently at the epoch where $\rho_m \sim \rho_\Lambda \gg \rho_\gamma$. At sufficiently late times, all the densities except the cosmological constant will have decayed away, leaving us with a universe dominated by cosmological constant, which, as we just learned, corresponds to an exponential expansion $a \propto \exp\left(\sqrt{\frac{\Lambda}{3}}t\right)$. Going into the past, we have a growing energy densities of matter ($\rho_m \propto a^{-3}$) and radiation ($\rho_\gamma \propto a^{-4}$). Matter soon largely dominates over the cosmological constant, resulting in a matter-dominated universe with a scale factor varying as $a \propto t^{2/3}$. Going further into the past, radiation eventually overtakes matter, leading to a radiation-domination epoch with $a \propto t^{1/2}$.

> Summary:
> As the universe expand, different energy components scale differently and, consequently, dominate the universe at different times. In the ΛCDM model, the universe started in a radiation-domination stage, then switched into an intermediary matter-domination stage, before finally entering a never-ending cosmological-constant-domination stage.

II Thermal History

review of equilibrium statistical physics; Big-bang theory and cosmic microwave background

5.2 Review of equilibrium statistical physics

With the knowledge we have so far, we can extrapolate back into the history of the universe until the temperature becomes high enough that interactions allow significant interchanges among the energy components of the universe. Probing back further would require assumptions about the particle interactions and the nature of physical laws themselves. Nevertheless, this apparent drawback, when taken to the extremity, turns into an advantage. As particles interact more strongly and rapidly, they are more likely to achieve thermal equilibrium. Thanks to that, we can use the powerful concept of thermal equilibrium, which allows us to describe a huge system, i.e. the universe, with only a few thermodynamical parameters such as temperature and chemical potentials. It is remarkable how much further we can go when armed with statistical mechanics.

A system in thermal equilibrium can be characterized by a density matrix. In quantum physics, the density matrix is given by

$$\hat{\rho} = \frac{1}{Z}\exp\left(-\frac{\hat{H}}{T} + \mu_i \hat{Q}_i\right) \tag{5.22}$$

where Z, $\hat{H}$, T, μ_i, and Q_i are respectively the partition function, Hamiltonian, temperature, chemical potentials, and conserved numbers (e.g. baryon number, lepton number) of the system. The density matrix satisfies $\mathrm{Tr}\,\hat{\rho} = 1$ and conserved numbers satisfy $[\hat{Q}_i, \hat{H}] = 0$, so that they are conserved $-\frac{\hbar}{i}\frac{d}{dt}\hat{Q}_i = [\hat{H}, \hat{Q}_i] = 0$. Classically, the density matrix is given by

$$\rho = \frac{1}{Z}\exp\left[-\frac{H(p,q)}{T} + \mu_i Q_i(p,q)\right] \tag{5.23}$$

with $\{H, Q_i\} = 0$.

In many cases, the full-blown treatment using the density matrix is not needed. A system can often be accurately described as a collection of non-interacting particles. The number density N and energy density ρ of free particles in thermal equilibrium are given by

$$N = \frac{g}{(2\pi)^3}\int d^3\mathbf{p}\, n(\mathbf{p}) \tag{5.24}$$

$$\rho = \frac{g}{(2\pi)^3}\int d^3\mathbf{p}\, E(\mathbf{p}) n(\mathbf{p}) \tag{5.25}$$

where g is the number of degrees of freedom, $E(\mathbf{p})$ is the energy of a particle with momentum $\mathbf{p}$, and $n(\mathbf{p})$ is the distribution function in thermal equilibrium, which is given by

$$n_{B/F}(\mathbf{p}) = \frac{1}{\exp\left(\frac{E(\mathbf{p})-\mu}{T}\right) \mp 1} \tag{5.26}$$

For fermions we pick the positive sign and for bosons we pick the negative sign. In the relativistic limit $T \gg m$, the integrals in (5.24) and (5.25) can be calculated explicitly. The results are

$$\rho = \begin{cases} \dfrac{\pi^2}{30} g T^4 & \text{boson} \\ \dfrac{7}{8}\dfrac{\pi^2}{30} g T^4 & \text{fermion} \end{cases} \tag{5.27}$$

$$N = \begin{cases} \dfrac{\zeta(3)}{\pi^2} g T^3 & \text{boson} \\ \dfrac{3}{4}\dfrac{\zeta(3)}{\pi^2} g T^3 & \text{fermion} \end{cases} \tag{5.28}$$

where $\zeta(x)$ is the Riemann zeta function. If both bosons and fermions are present, the total energy density can be written compactly as

$$\rho = \frac{\pi^2}{30} g_* T^4 \tag{5.29}$$

where

$$g_* \equiv \sum_{\text{bosons}} + \frac{7}{8} \sum_{\text{fermions}} \tag{5.30}$$

is the effective number of relativistic degrees of freedom. The sums above are taken over species that are relativistic at temperature T. The entropy per unit volume s, can be expressed in terms of the distribution function as

$$s_{B/F} = -\int \frac{d^3\mathbf{p}}{(2\pi)^3} \left[n_{B/F}(\mathbf{p}) \log n_{B/F}(\mathbf{p}) \mp (1 \pm n_{B/F}(\mathbf{p})) \log \left(1 \pm n_{B/F}(\mathbf{p})\right) \right] \tag{5.31}$$

In the relativistic limit, $n_{B/F} << 1$ and we get a common formula

$$s = -\int \frac{d^3\mathbf{p}}{(2\pi)^3} n(\mathbf{p}) \left(1 - \log n(\mathbf{p})\right) \tag{5.32}$$

Alternatively, we can obtain it via the following thermodynamic argument. In the absence of chemical potential, we have

$$U = TS - pV \tag{5.33}$$

Taking the volume V to be a comoving volume, we can define the entropy density as

$$s \equiv \frac{\partial S}{\partial V} = \frac{\rho + p}{T} \tag{5.34}$$

Using (5.27) and the equation of state for radiation $p = \rho/3$, we can calculate the entropy density s in the relativistic limit

$$s = \frac{2\pi^2}{45} g_* T^3 \tag{5.35}$$

In the non-relativistic limit, on the other hand, the number density n, energy density ρ, and pressure p are given by

$$N = g \left(\frac{mT}{2\pi}\right)^{3/2} \exp\left[-\frac{(m-\mu)}{T}\right] \tag{5.36}$$

$$\rho = gmN \tag{5.37}$$

$$p = gNT \ll \rho \tag{5.38}$$

for both bosons and fermions.

5.3 Big-Bang theory and the cosmic microwave background

In the ΛCDM model, the universe at its earliest stage is dominated by radiation. According to (5.12), the energy density in a radiation dominated universe becomes singular as $t \to 0$. What happens at the singularity is not yet answerable by physics. The known laws of physics are expected to break down at least around the Planck scale $E \sim M_{\text{pl}} = 1.22 \times 10^{19}$ GeV; to describe physics beyond the Planck scale, a theory of quantum gravity is needed. Despite that, we can at least deduce that close to the singularity, but before the known laws of physics break down, the universe was filled with densely-packed and rapidly-interacting particles, which are presumably in a thermal equilibrium with a very high temperature. This suggests that the universe began in an extremely hot and dense state that one could describe as a "bang," hence the name Big-Bang theory.

One prediction that comes out of this reasoning is that of the cosmic microwave background (CMB), which is commonly attributed to George Gamow. The Big-Bang theory says that the universe was initially hot and dense, full of photons, electrons, ions, etc. It then cools down as it expands. At temperatures of around the Rydberg energy 13.6 eV, i.e. typical binding energy of atoms, ions and electrons start to bind together to form atoms, and consequently the universe became transparent to radiation. Following this reasoning, we expect the present universe to be populated by relic photons that were once in thermal equilibrium, but have decoupled, lingering around freely as a background, the CMB, that we can observe. In 1965, the CMB was discovered accidentally by Penzias and Wilson at temperature of around 4 K (more precise measurements narrowed it down to around 2.73 K).

Let us now try to estimate the current CMB temperature with the tools we have in our hands. (5.29) allows us to express the Hubble parameter in the radiation-domination epoch as

$$H = \left(\frac{8\pi G}{3}\rho_{\text{rad}}\right)^{1/2} = \frac{T^2}{M_0} \tag{5.39}$$

where we have defined

$$M_0 \equiv \left(\frac{45}{4\pi^3 g_* G}\right)^{1/2} = \frac{M_{\text{pl}}}{1.66 g_*^{1/2}} \tag{5.40}$$

Combining (5.18) and (5.39), we obtain a relation between the time t and temperature T

$$t = \frac{M_0}{2T^2} = 0.301 g_*^{-1/2} \frac{M_{\text{pl}}}{T^2} \sim 1\ \text{s} \times \left(\frac{T}{1\ \text{MeV}}\right)^{-2} \tag{5.41}$$

This is one of the most useful equations of the Big-Bang theory. The temperature of the CMB photons can be estimated (very roughly) by plugging the age of the universe into t. Plugging $t \sim 10$ billion years gives $T \sim 1$ meV $\sim$ 10K, which is roughly in accordance with the measured value (2.73 K).

Another prediction of the big bang theory is about the abundances of light elements in the universe, such as Li^7, D, He^3, He^4, etc which cannot be explained by thermodynamic cooking in stars alone. A process known as the Big Bang Nucleosynthesis (BBN) could explain their abundances quite accurately. We will learn about BBN later in the course.

5.4 Aside: g^* and the Standard Model

Let us count the number of degrees of freedom in the standard model.

Scalar sector

Higgs field h: one degree of freedom. Goldstone bosons $\{\theta_1, \theta_2, \theta_3\}$ (which disappear when the W and Z bosons become massive): three times one degree of freedom.

Fermion sector

The following quarks carry one fermionic degree of freedom each:

$$\{u_L, u_R, d_L, d_R, c_L, c_R, s_L, s_R, t_L, t_R, b_L, b_R\}. \tag{5.42}$$

In the Standard Model, all these quarks come in three colours. Also, next to every quark there exists the corresponding antiquark. So we end up with $12 \cdot 3 \cdot 2 = 72$ fermionic degrees of freedom at the quark side.

Next, there are the leptons. The following leptons carry one fermionic degree of freedom each:

$$\{e_L, e_R, \nu_L^e, \mu_L, \mu_R, \nu_L^\mu, \tau_L, \tau_R, \nu_R^\tau\}. \tag{5.43}$$

(In the Standard Model, there are no righthanded neutrinos.) Leptons do not feel color ($SU(3)$) charge. There is an antilepton next to every lepton though. So we count $9 \cdot 2 = 18$ fermionic degrees of freedom at the lepton side.

Vector sector

Massless vectors carry two degrees of freedom, for their two polarizations. Massive vectors have three polarizations, so three degrees of freedom. After electroweak symmetry breaking, the Standard Model contains nine massless vector (the photon plus eight gluons) and three massive vectors ($\{W^+, W^-, Z^0\}$).

Total

$$\begin{aligned} g_*^{\text{SM}} &= \sum_{\text{bosons}} g_*(\text{boson}) + \frac{7}{8} \cdot \sum_{\text{fermions}} g_*(\text{fermion}) \\ &= 1 + 9 \cdot 2 + 3 \cdot 3 + \frac{7}{8}(72 + 18) \\ &= 106.75. \end{aligned} \tag{5.44}$$

> Summary:
> The big bang theory is based on the simple deduction that since the universe is expanding (or shrinking as we go backwards in time) it must be very dense and hot in the past. One of the predictions of the big bang theory is the presence of the cosmic microwave background (CMB), freely-lingering relic photons that were once in thermal equilibrium but decoupled at some point.

Lecture 6

kinetic equations; relaxation-time approximation; freeze in and freeze out.

6.1 Particle kinetics in an expanding universe

To a good approximation, we can describe the universe as a homogeneous and isotropic gas of particles with a distribution function $n(\mathbf{p},\mathbf{x},t) = n(|\mathbf{p}|,t)$. In the absence of inter-particle interactions, the momentum simply scales as $\mathbf{p} \propto a^{-1}$ as the universe expands (this was derived in the case of a photon in lecture 3), and so the distribution function evolves as

$$n(\mathbf{p},t) = n(\mathbf{p}_0,t_0) = n_0\left(\mathbf{p}\frac{a}{a_0}\right) \tag{6.1}$$

where n_0 is the distribution function at $t = t_0$. In words, the distribution function is unchanged apart from a rescaling in the momentum to account for the fact that particles whose momentum is $\mathbf{p}$ at an arbitrary time t have momentum $\mathbf{p}_0 = \mathbf{p}a/a_0$ at time $t = t_0$. The equation describing the evolution of n, called kinetic equation, can be derived by relating the partial derivative

$$\frac{\partial n}{\partial t} = \frac{\partial n_0}{\partial \mathbf{p}_0}\frac{\partial \mathbf{p}_0}{\partial t} = \frac{\partial n_0}{\partial \mathbf{p}_0}\mathbf{p}\frac{\dot{a}(t)}{\dot{a}(t_0)} \tag{6.2}$$

where $\mathbf{p}_0 = \mathbf{p}a/a_0$ and a partial derivative with respect to a vector $\mathbf{V}$ is defined as $\partial/\partial\mathbf{V} = \partial/\partial V_x\hat{x} + \partial/\partial V_y\hat{y} + \partial/\partial V_z\hat{z}$, with the partial derivative

$$\frac{\partial n}{\partial \mathbf{p}} = \frac{\partial n_0}{\partial \mathbf{p}_0}\frac{\partial \mathbf{p}_0}{\partial \mathbf{p}} = \frac{\partial n_0}{\partial \mathbf{p}_0}\frac{a}{a_0} \tag{6.3}$$

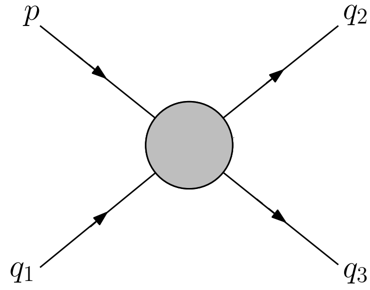
Substituting the $\partial n_0/\partial\mathbf{p}_0$ in (6.3) into the one in (6.2), gives us the kinetic equation

$$\frac{\partial n}{\partial t} - H\mathbf{p}\frac{\partial n}{\partial \mathbf{p}} = 0 \tag{6.4}$$

A similar equation for the number density $N = \int \frac{d^3p}{(2\pi)^3}n(\mathbf{p})$ can be obtained by taking the integral $\int \frac{d^3p}{(2\pi)^3}$ of the above equation

$$\frac{dN}{dt} - \int \frac{d^3p}{(2\pi)^3}H\mathbf{p}\frac{\partial N}{\partial \mathbf{p}} = 0 \tag{6.5}$$

Figure 6.1: Schematic of a $2 \leftrightarrow 2$ collision.



The integral term can be simplified via integration by parts for each of the momentum component, e.g. $\int d^3p p_x \partial n/\partial p_x = -\int d^3p \partial p_x/\partial p_x n = -N$, where we have used $n(p) \to 0$ for $p \to \infty$. Doing so, yields

$$\frac{dN}{dt} + 3HN = 0 \tag{6.6}$$

It is easy to see that the solution corresponds to the scaling law $N \propto a^{-3}$.

The kinetic equation (6.4) was derived by neglecting the interaction among particles. To account for collisions, we can simply add a collision integral term I_{col} so that it becomes

$$\frac{\partial n}{\partial t} - H\mathbf{p}\frac{\partial n}{\partial \mathbf{p}} = I_{\text{col}} \tag{6.7}$$

which is known as the *Boltzmann equation*. For example, to account for $2 \leftrightarrow 2$ scatterings (Figure. 6.1), we include

$$I_{\text{col}} = -\frac{1}{2p^0}\int \frac{d^3q_1}{(2\pi)^3 2q_1^0}\frac{d^3q_2}{(2\pi)^3 2q_2^0}\frac{d^3q_3}{(2\pi)^3 2q_3^0}(2\pi)^4\delta^4(p+q_1-q_2-q_3)\,|M_{fi}|^2$$
$$\times\{n(p)n(q_1)[1\pm n(q_2)][1\pm n(q_3)] - n(q_3)n(q_2)[1\pm n(p)][1\pm n(q)]\} \tag{6.8}$$

The delta function is there to ensure 4-momentum conservation, $|M_{fi}|^2$ is the matrix element of the process, and the first and second term in the curly bracket accounts for the phase space of the initial and final states respectively with Bose-enhancements (positive signs) or Fermi-suppressions (negative signs) included. In general, we could have other types of collisions, such as $1 \leftrightarrow 2$ (e.g. $W^+ \leftrightarrow \mu^+\nu_\mu$), $1 \leftrightarrow 3$ (e.g. $\mu^+ \leftrightarrow e^+\nu_e\bar{\nu}_\mu$), $2 \leftrightarrow 3$ (e.g. $e^+e^- \leftrightarrow 3\gamma$), etc, and the I_{col} is given by the sum of the contribution of each collision-type. The Boltzmann equation (6.7) is an integro-differential equation, which is very difficult to solve in general. In many cases of interest, however, the universe is close to being in thermal equilibrium, and an approximation dubbed the relaxation-time approximation can be used. This approximation is motivated by the expectation that the effect of particle collisions, which manifests itself through the collision term I_{col}, is driving the system towards thermal equilibrium, whereupon $I_{\text{col}} = 0$. Inspired by that, in the relaxation-time approximation we assume that $I_{\text{col}} \propto -(n - n_{\text{eq}})$ and write the Boltzmann equation as

$$\frac{\partial n}{\partial t} - H\mathbf{p}\frac{\partial n}{\partial \mathbf{p}} = -\Gamma(n - n_{\text{eq}}) \tag{6.9}$$

where Γ, the rate of reaction, is given by the inverse of the mean free time τ of the particle collisions

$$\Gamma = \tau^{-1} = \left(\left\langle\frac{\lambda}{v}\right\rangle\right)^{-1} = \langle\sigma N v\rangle \tag{6.10}$$

where v is the relative velocity of the colliding particles, $\lambda = (\sigma n)^{-1}$ is the mean free-path of particles with cross-section σ. Notice that, in the absence of expansion (6.9) correctly predicts that n tends to relax to n_{eq}.

The Boltzmann equation for the number density n in the relaxation-time approximation can be found by taking the integral $\int \frac{d^3p}{(2\pi)^3}$ of (6.9). The left hand side was found in (6.6) and the right hand side is simply given by $-\Gamma(N - N_{\text{eq}})$. Putting them together, we have

$$\frac{dN}{dt} + 3HN = -\Gamma(N - N_{\text{eq}}) \tag{6.11}$$

6.2 Freeze in and freeze out

We found in (6.11) that the evolution of the number density N of particles in an expanding universe is driven by two effects: the diluting effect of the expansion, whose strength is characterized by the Hubble parameter H, and the thermalizing effect of particle collisions, whose strength is characterized by the reaction rate Γ. The temperature T_* at which

$$\Gamma(T_*) = H(T_*) \tag{6.12}$$

marks the transition between two qualitatively different behaviours of the number density N. In one extreme $\Gamma(T) \gg H(T)$, the expansion of the universe is negligible and thermal equilibrium can be achieved without hurdle. In the other extreme $\Gamma(T) \ll H(T)$, interactions happen so rarely that the number of particles in a comoving volume essentially freezes and hence the number density simply scales as $N \propto a^3$. These behaviours are summarized below

$$N \begin{cases} \approx N_{\text{eq}}(T), & \Gamma(T) \gg H(T) \\ \sim N_{\text{eq}}(T_*) \left(\dfrac{a(T_*)}{a(T)} \right)^3, & \Gamma(T) \ll H(T) \end{cases} \tag{6.13}$$

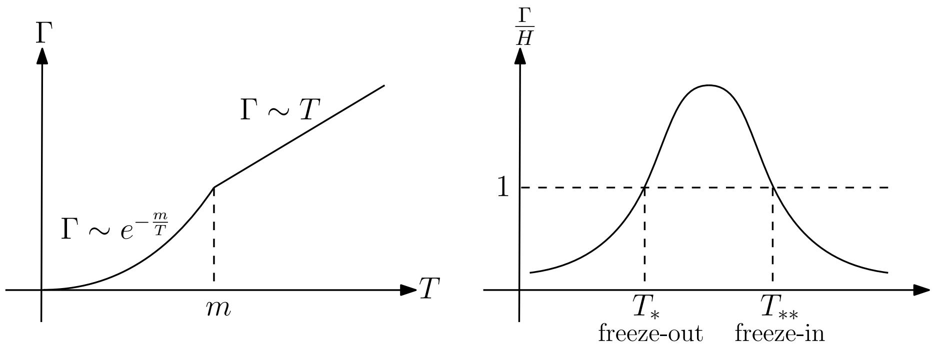
In the expression for $\Gamma(T) \ll H(T)$, we have picked the point $T = T_*$ as the scaling reference for N. Since at that point $\Gamma(T_*) = H(T_*)$, and not $\Gamma(T_*) \ll H(T_*)$, the expression is only correct marginally, up to an O(1) factor. T_* is sometimes called *freeze-in* temperature and sometimes called *freeze-out* temperature, depending on how the ratio Γ/H changes with time. Freeze in happens when the system starts in the out-of-equilibrium regime, $\Gamma(T)/H(T) \ll 1$, and moves towards the $\Gamma(T)/H(T) \gg 1$ regime, getting thermalized in the process. Freeze out is the opposite process occurring when an initially thermalized system in the $\Gamma(T)/H(T) \gg 1$ regime moves towards the $\Gamma(T)/H(T) \ll 1$ regime, getting driven out of equilibrium in the process. We will see these processes more explicitly when we study the example below. Before going any further, we would like to point out two caveats of (6.12) and (6.13). First, $\Gamma(T)$ here refers to only reactions / collisions that change the number of particles under consideration. Second, (6.13) is not applicable for unstable, decaying particles.

To get a grasp of the concept of freeze in and freeze out, let us take a look at a gas of electron and positron, interacting mainly via the process $e^+e^- \leftrightarrow \gamma\gamma$[1], in the radiation domination epoch. The cross section of the process can be calculated in the center of mass frame to be

$$\sigma \approx \begin{cases} \dfrac{1}{2v} \pi r_e^2, & v \ll 1 \\ \dfrac{m_e^2}{E^2} \pi r_e^2 \left(\log \dfrac{4E^2}{m_e^2} - 1 \right), & v \approx 1 \end{cases} \tag{6.14}$$

[1] There are of course other processes such as $e^-\gamma \leftrightarrow e^-\gamma$, $e^+\gamma \leftrightarrow e^+\gamma$, $e^-e^+ \leftrightarrow e^-e^+$, etc, but those processes merely interchange the momenta of the particles involved without changing their particle numbers and therefore irrelevant as far as freeze-in or freeze-out processes are concerned.

Figure 6.2: Temperature dependence of the rate of interaction Γ (left) and its relative size compared to the Hubble parameter Γ/H (right) for electron-positron-photon system in the radiation domination epoch.



where $r_e = \alpha/m_e$ with $\alpha = 1/137$ being the fine-structure constant and E is the total energy of the colliding particles in the center of mass frame. The derivation of (6.14) is beyond the scope of these lectures. Nevertheless, one could at least guess its $\sigma \propto E^{-2}$ dependence, which is valid in both non-relativistic and relativistic case, by dimensional considerations. Dismissing the logarithmic term, which is subleading and for simplicity, the rate of reaction Γ for the cross-section (6.14) can be computed as

$$\Gamma = \langle \sigma N v \rangle \approx \begin{cases} \dfrac{\pi r_e^2}{2} 2 \left(\dfrac{m_e T}{2\pi}\right)^{3/2} e^{-m_e/T}, & T \lesssim m_e \\ \pi\alpha^2 \dfrac{1}{6}\dfrac{\zeta(3)}{\pi^2} T, & T \gtrsim m_e \end{cases} \tag{6.15}$$

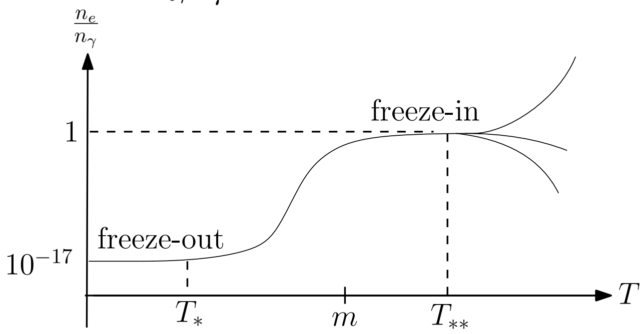
where we have used (5.36) in the $T \lesssim m_e$ case, and (5.28) and $E \sim 3T$ in the $T \gtrsim m_e$ case. The qualitative behaviour of $\Gamma(T)$ is depicted in Figure 6.2; it is large at high T and small at low T, which is in accordance with our intuition: higher temperature implies higher thermal equilibrium density, which, in turn, implies that the particles interact more frequently, hence the higher rate Γ. In the radiation epoch, the Hubble parameter $H(T)$ is given by (5.39), so

$$\frac{\Gamma(T)}{H(T)} = \begin{cases} \pi r_e^2 \left(\dfrac{m_e T}{2\pi}\right)^{3/2} e^{-m_e/T} \dfrac{M_0}{T^2}, & T \lesssim m_e \\ \pi\alpha^2 \dfrac{1}{6}\dfrac{\zeta(3)}{\pi^2} T \dfrac{M_0}{T^2}, & T \gtrsim m_e \end{cases} \tag{6.16}$$

As can be seen in Figure 6.2, $\Gamma(T)/H(T)$ hits 1 twice, at $T = T_{**}$ (freeze in) and $T = T_*$ (freeze out). These crossing points can be found by equating the above expression to 1. The freeze-in temperature T_{**}, which lies in the $T \gtrsim m_e$ regime, is easier to calculate

$$T_{**} = \alpha^2 \frac{\zeta(3)}{6\pi} M_0 \sim 10^{14} \text{ Gev} \tag{6.17}$$

Figure 6.3: The behaviour of N_e/N_γ during the freeze-in and freeze-out process. The three different curves on the right part of the graph illustrate the fact that the final, freeze-out value of N_e/N_γ does not depend on the initial value of N_e/N_γ before freeze-in.



Calculating the freeze-out temperature T_*, which lies in the $T \lesssim m_e$ regime, is more involved

$$\begin{aligned} \pi r_e^2 \left(\frac{m_e T_*}{2\pi}\right)^{3/2} e^{-m_e/T_*} \frac{M_0}{T_*^2} &= 1 \\ \pi r_e^2 \left(\frac{m_e^2}{2\pi x}\right)^{3/2} e^{-x} \frac{M_0}{m_e^2} x^2 &= 1 \\ x^{1/2} e^{-x} &\approx \frac{2\sqrt{2\pi}}{\alpha^2} \frac{m_e}{M_0} \\ \frac{1}{2}\log x - x &\approx -39 \end{aligned} \tag{6.18}$$

From the second to third line, $x \equiv m_e/T_*$ was introduced; from the third to fourth line $r_e = \alpha/m_e$ was used; in the last step, we plugged in the numerical values of the quantities. The last equation can be solved iteratively or perturbatively, treating the log term as a small correction

$$\frac{m_e}{T_*} \equiv x \approx 39 + \frac{1}{2}\log 39 \approx 43 \tag{6.19}$$

After the freeze out, the comoving number density freezes at around its value at $T = T_*$.

In this example, the photons are always in thermal equilibrium and consequently its number density simply scales as $N_\gamma \propto a^{-3}$. This means that we can use the ratio N_e/N_γ as a representative of the comoving number density of the electrons. The behaviour of N_e/N_γ during the freeze in and freeze out is shown in Figure. 6.3. As shown in the figure, after freeze out ($T < T_*$) N_e/N_γ freezes at its value at $T = T_*$

$$\left.\frac{N_e}{N_\gamma}\right|_{T \lesssim T_*} \sim \left.\frac{N_e}{N_\gamma}\right|_{T = T_*} \approx \left(\frac{43}{2\pi}\right)^{3/2} e^{-43} \frac{\pi^2}{\zeta(3)} \sim 10^{-17} \tag{6.20}$$

Note that the process of freeze in, which is nothing but thermalization, erases any information about the initial conditions of electrons in the universe. Consequently, the freeze-out density of N_e/N_γ does not depend on the initial state of the electrons at $T > T_*$.

Summary:
The universe can be approximated as a gas of particles whose distribution function (and number density) evolves according to the kinetic equation / Boltzmann equation (6.7). In the relaxation-time approximation, the Boltzmann equation for the number density simplifies to (6.11). The behaviour of the number density is qualitatively different depending on the relative size between

the interaction rate Γ and the Hubble parameter H. If $\Gamma \gg H$, the particles are in a thermal equilibrium. If $\Gamma \ll H$, the number density N simply scales as $N \propto a^{-3}$. When $\Gamma \sim H$, there is a transition from one behaviour to the other. If the transition goes from $\Gamma \ll H$ (out of thermal equilibrium) to $\Gamma \gg H$ (in thermal equilibrium), the process is called freeze in and if it is the other way around it is called freeze out.

Lecture 7

decoupling of photons; entropy conservation;

7.1 Decoupling of photons

At some temperature far above 13.6 eV, the Rydberg energy (Hydrogen's binding energy), the universe was mainly composed of tightly-coupled p, e^-, γ, and H plasma soup. At this stage, due to the presence of charged particles, the mean free path of photons was much shorter than the Hubble radius. At temperatures far below the Rydberg energy, it became thermodynamically favourable for the e^-'s and p's to combine into neutral atoms. As a result, the mean free path of photons became much larger than the Hubble radius and the photons essentially decoupled from the rest of the soup; these decoupled photons are the ones that we observe today as the CMB. Here, we are interested in studying the in-between process: how does the plasma soup behave as the temperature of the universe crosses the Rydberg energy?

Before the decoupling of photons, the main reaction holding photons in thermal equilibrium is $e^- + \mathrm{p} \leftrightarrow \mathrm{H} + \gamma$. When the species involved are in thermal equilibrium, they obey the Saha's equation

$$\frac{n_e n_p}{n_{\mathrm{H}}} = \left(\frac{m_e T}{2\pi}\right)^{3/2} e^{-I/T} \tag{7.1}$$

where $I = m_p + m_e - m_{\mathrm{H}} = 13.6\ \mathrm{eV} = 1.58 \times 10^5$ K is the Rydberg energy.

Remark: derivation of the Saha's equation (7.1).
The number density of photons in thermal equilibrium is given by (5.28). Since we are interested in studying a low-energy process occurring at $T \sim 13.6$ eV, the species other than photon can be assumed to be non-relativistic, with their number densities given by (5.36)

$$n_e = g_e \left(\frac{m_e T}{2\pi}\right)^{3/2} \exp\left[-\frac{(m_e - \mu_e)}{T}\right] \tag{7.2}$$

$$n_p = g_p \left(\frac{m_p T}{2\pi}\right)^{3/2} \exp\left[-\frac{(m_p - \mu_p)}{T}\right] \tag{7.3}$$

$$n_{\mathrm{H}} = g_{\mathrm{H}} \left(\frac{m_{\mathrm{H}} T}{2\pi}\right)^{3/2} \exp\left[-\frac{(m_{\mathrm{H}} - \mu_{\mathrm{H}})}{T}\right] \tag{7.4}$$

where $g_e = g_p = 2$ and $g_{\rm H} = 2\times 2 = 4$. Noting that $\mu_\gamma = 0$, when the plasma is in a chemical equlibrium the chemical potentials satisfy

$$\mu_p + \mu_e = \mu_{\rm H} \tag{7.5}$$

The desired Saha's equation (7.1) is obtained by multiplying (7.2) with (7.3) and dividing the result with (7.4).

Assuming that the universe was neutral as a whole, we can set $n_e = n_p$ so that the Saha's equation reduces to

$$n_e \approx n_{\rm H}^{1/2}\left(\frac{m_e T}{2\pi}\right)^{3/4} e^{-\frac{I}{2T}} \tag{7.6}$$

Big Bang Nucleosynthesis considerations, which will be studied later in the lectures, constraint the value of the baryon to entropy (photon) ratio quite precisely to around $\eta \equiv n_B/n_\gamma = 6.15\times 10^{-10}$. Thus, assuming that Hydrogen makes up most of the baryon density, we have

$$n_{\rm H} \approx n_B = \eta n_\gamma = \eta\frac{\zeta(3)}{\pi^2}2T^3 \tag{7.7}$$

where (5.28) was used in the last equality. The reaction rate for the process $e^- + {\rm p} \leftrightarrow {\rm H} + \gamma$ is given by

$$\Gamma = \langle \sigma_{\gamma e} n_e v\rangle + \langle \sigma_{\gamma p} n_p v\rangle \tag{7.8}$$

with

$$\sigma_{\gamma e} = \frac{8\pi}{3}r_e^2 = \frac{8\pi\alpha^2}{3m_e^2} \tag{7.9}$$

$$\sigma_{\gamma p} = \frac{8\pi}{3}r_p^2 = \frac{8\pi\alpha^2}{3m_p^2} \tag{7.10}$$

which, again, can be deduced by dimensional arguments. In (7.8), we have neglected the contribution from γ-H scattering, which is presumably small relative to the others as Hydrogen is neutral. Owing to $m_e \ll m_p$, which implies that $\sigma_{\gamma e} \gg \sigma_{\gamma p}$, and the neutrality of the universe $n_e = n_p$ we can neglect the photon-proton contribution to the reaction rate, leaving us with

$$\Gamma = \langle \sigma_{\gamma e} n_e v\rangle = \frac{8\pi\alpha^2}{3m_e^2}\sqrt{\eta\frac{\zeta(3)}{\pi^2}2T^3}\left(\frac{m_e T}{2\pi}\right)^{3/4} e^{-\frac{I}{2T}} \tag{7.11}$$

where we have used (7.6), (7.7), (7.9), and $v \sim 1$ (since photons are involved) in arriving at the last expression. As the temperature decrease due to the expansion, $\Gamma(T)$ gets increasingly Boltzmann suppressed. When $\Gamma(T) = H(T)$, the plasma looks essentially transparent to the photons and photon decoupling occurs. Equating the reaction rate above with the Hubble parameter (5.39) gives

$$B\frac{T^{9/4}}{m_e^{5/4}}e^{-\frac{I}{2T}} = \frac{T^2}{M_0} \tag{7.12}$$

where we have encapsulated some of the constants in

$$B = \frac{8\pi\alpha^2}{3}\eta^{1/2}\left(\frac{2\zeta(3)}{\pi^2}\right)^{1/2}\left(\frac{1}{2\pi}\right)^{3/4} \tag{7.13}$$

After defining $x = I/2T$, (7.12) simplifies to

$$\frac{1}{x^{1/4}}e^{-x} = \frac{1}{B}\frac{2^{5/4}m_e^{5/4}}{I^{1/4}M_0} = 1.24 \times 10^{-12} \tag{7.14}$$

which can be solved numerically to give

$$x = 26.6 \tag{7.15}$$

or

$$T_{\text{dec}} = \sqrt{\sigma_{\gamma e} n_e M_0} = 3000 \text{ K} = 0.25 \text{ eV} \tag{7.16}$$

Unsurprisingly, T_{dec} is of the order of the Rydberg energy, where we would expect the Boltzmann suppression to kick-in. The decoupling temperature is sometimes referred to as the recombination temperature, reflecting the fact that protons and electrons combine together to form neutral atoms at that temperature. It corresponds to the redshift

$$z_{\text{dec}} = \frac{T_{\text{dec}}}{T_0} - 1 \approx 1100 \tag{7.17}$$

$T_0 = 2.73$ K is the current CMB temperature. We have made various assumptions in deriving this result, including but not limited to not accounting for the Hydrogen-photon cross-section and the full spectrum of bound states other than the ground state. However, the result turns out to be pretty accurate. Incidentally, z_{dec} happens to have the same order of magnitude as the redshift of matter-radiation equality, $z_{\text{eq}} \approx 3000$. Whether this is a pure coincidence or whether there exists some explanation connecting the two remains to be investigated.

> Summary:
> The photons that we see today as the CMB decoupled from the rest of the universe at the temperature of $T_{\text{dec}} = 0.25$ eV corresponding to the redshift $z_{\text{dec}} \approx 1100$.

Lecture 8

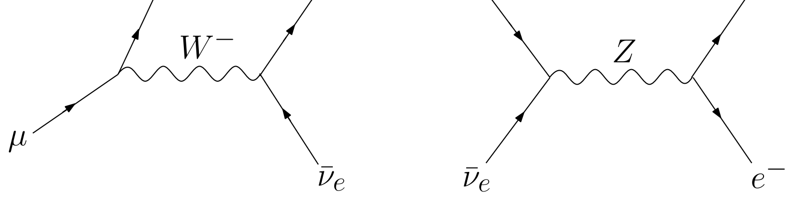
neutrino-decoupling temperature; present temperature of relic neutrinos; cosmological constraints on neutrino masses

8.1 Freeze-out and present temperature of neutrinos

Neutrinos are weakly interacting particles which come in three flavours: electron neutrino ν_e, muon neutrino ν_μ, and tau neutrino ν_τ. Each neutrino type carries a lepton number $L = 1$ and a conserved number of value 1 associated with its type, e.g. a muon neutrino has muon number $n_\mu = 1$ and so forth. This ensures that different neutrino types do not mix in interactions (but they may mix when they propagate). Neutrinos most probably have masses, though they must be very small ($\sim$ eV or less). One of the strongest bounds on the neutrino masses comes from cosmology. Later in the lectures, we will derive this cosmological bound. Neutrinos are spin-1/2 particles, but unlike, say, electron which has two degrees of freedom (corresponding to left-handed and right-handed helicity), they have only a single degree of freedom, $g = 1$. The reason being that all neutrinos that we have observed so far have left-handed helicity, i.e. their spins are antiparallel to their momentum (similarly, the antineutrinos that we have observed are all right-handed). The possibility that there exist right-handed neutrinos is not entirely ruled out. However, if they do exist, they must interact extremely weakly with the particles that we have observed.

Apart from the CMB photons, the present universe is also filled with relic neutrinos that were once in thermal equilibrium but decoupled at some point. Let us estimate the present temperature of these relic neutrinos. We will see that it turns out to be different from that of the CMB. For the current purpose, we will assume that the neutrinos are massless. Before their decoupling, the neutrinos were held in thermal equilibrium mainly by the charged-current and neutral-current interactions, both of which are depicted in Figure 8.1. The cross-section for these interactions are

Figure 8.1: Charged-current (left) and neutral-current (right) interactions involving neutrinos.



roughly

$$\langle \sigma_W v\rangle \sim G_F^2 E^2 \tag{8.1}$$

with $G_F = g^2/M_W^2 \sim 10^{-5}$ GeV^{-2}. For $E \sim 1$ MeV, we have $\sigma_W \sim 10^{-16}$ GeV^{-2}, which is tiny in comparison to typical cross section for electromagnetic interactions at the same energy $\sigma_{\mathrm{EM}} \sim \alpha^2/E^2 \sim 10^2$ GeV^{-2}. Neutrino decoupling happens when the rate of reaction

$$\Gamma_W = \langle \sigma_W n v\rangle \sim \left(G_F^2 T^2\right)\left(T^3\right)(1) = G_F^2 T^5 \tag{8.2}$$

is equal to the Hubble parameter (5.39). This happens at

$$T_* \sim \left(G_F^2 M_0\right)^{-1/3} \sim 2\ \mathrm{MeV} \tag{8.3}$$

which is incidentally pretty close to the correct value we would obtain if the $O(1)$ factors are taken into account. We found that T_* is much larger than the electron mass but much smaller than the masses of all particles other than photons and neutrinos. It means that at $T \approx T_*$ the universe is populated by a thermal equilibrium mixture of e^+, e^-, γ, ν_e, ν_μ, and ν_τ.

8.2 Boltzmann H-theorem

Now, before going on and estimating the temperature of relic neutrinos, we want to consider in more details the behaviour of entropy in an expanding universe. Without expansion, the total entropy is conserve. As long as we have a "close-too-adiabatic" expansion, we would expect the entropy to be closed to conserved. And this is actually the case, as we will see in this section and the next. In this section, we will look at Boltzmann H-theorem, which shows that the entropy defined from the number density cannot decrease over time. We begin from the definition of the entropy density s in terms of the occupation number $n(p)$:

$$s = -\int \frac{d^3 p}{(2\pi)^3}\, n(p)\log\left(n(p)/e\right). \tag{8.4}$$

(Check dimensions: at the left hand side we have one over a volume, so mass dimension 3, at the right hand side the only dimensionful quantity is d^3p, so also mass dimension 3.)

In thermal equilibrium we have (using the non-relativistic distribution for simplicity)

$$n(p) = e^{\frac{-E-\mu}{T}}, \tag{8.5}$$

but the relation in eq. 8.4 is also valid out of thermal equilibrium. Now, the Boltzmann theorem states that the dynamics of the macroscopic system is such that $\frac{ds}{dt} \geq 0$.

Boltzmann's proof comes from his kinetic equation. Take a simple reaction (elastic scattering)

$$p_1 + p_2 \leftrightarrow p_3 + p_4, \tag{8.6}$$

and ignore (just as in Boltzmann's time) all further subtleties like Fermi suppression, Bose enhancement, relativistic effects and the expansion of the universe. (Note that for fixed volumes, s is proportional to S.)

For the kinetic equation we then have ($\mathscr{D}p \equiv \frac{dp}{(2\pi)^3}$)

$$\begin{aligned}\frac{dn(p_1)}{dt} &= I_{\mathrm{coll}}(p_1)\\ &= \int \mathscr{D}p_2\mathscr{D}p_3\mathscr{D}p_4\,|\mathscr{M}_{\mathrm{f}\to\mathrm{i}}|^2\,\delta^{(4)}\left(p_1+p_2-p_3-p_4\right)\times\left[n(p_3)n(p_4)-n(p_1)n(p_2)\right].\end{aligned} \tag{8.7}$$

The intuitive understanding here is that when the densities of particles 3 and 4 are higher than those of particles 1 and 2 (“when the equilibrium in eq. 8.6 is shifted to the right"), the system responds by creating more particles 1 and 2, so then $\frac{dn(p_1)}{dt}$ is positive.

For the change of entropy density contained in this system we now have

$$\begin{aligned} \frac{ds}{dt} &= \frac{d}{dt}\left[-\int \frac{d^3p_1}{(2\pi)^3} n(p_1)\log\left(n(p_1)/e\right)\right] \\ &= -\int \mathscr{D}p_1 I_{\text{coll}}(p_1)\log\left(n(p_1)/e\right) - \int Dp_1 \dot{n}(p_1) \\ &= -\int \mathscr{D}p_1 I_{\text{coll}}(p_1)\log\left(n(p_1)/e\right), \end{aligned} \tag{8.8}$$

where we have dropped the second term because is is the time derivative of the total number of particles $\int Dp_1 n(p_1)$, which is conserved.

Now, replacing $\log\left(n(p_1)/e\right)$ by a general function $\Phi(p_1)$ for the moment, we have

$$\begin{aligned} I &\equiv \int \mathscr{D}p_1 I_{\text{coll}}(p_1)\Phi(p_1) \\ &= \int \mathscr{D}p_1\mathscr{D}p_2\mathscr{D}p_3\mathscr{D}p_4 \left|\mathscr{M}_{\text{f}\to\text{i}}\right|^2 \delta^{(4)}\left(p_1+p_2-p_3-p_4\right)\times\left[n(p_3)n(p_4)-n(p_1)n(p_2)\right]\Phi(p_1). \end{aligned} \tag{8.9}$$

Now we just change notation. When we set $p_1 \leftrightarrow p_2$ and $p_3 \leftrightarrow p_4$ (all momenta are integrated over, so we are free to exchange p_i's as we want) we get

$$I = \int \mathscr{D}p_1\mathscr{D}p_2\mathscr{D}p_3\mathscr{D}p_4 \left|\mathscr{M}_{\text{f}\to\text{i}}\right|^2 \delta^{(4)}\left(p_1+p_2-p_3-p_4\right)\times\left[n(p_3)n(p_4)-n(p_1)n(p_2)\right]\Phi(p_2). \tag{8.10}$$

Interchanging $p_1 \leftrightarrow p_3$ and $p_2 \leftrightarrow p_4$ in eq. 8.9 gives an additional minus sign:

$$I = -\int \mathscr{D}p_1\mathscr{D}p_2\mathscr{D}p_3\mathscr{D}p_4 \left|\mathscr{M}_{\text{f}\to\text{i}}\right|^2 \delta^{(4)}\left(p_1+p_2-p_3-p_4\right)\times\left[n(p_3)n(p_4)-n(p_1)n(p_2)\right]\Phi(p_3), \tag{8.11}$$

and the same happens when we interchange $p_1 \leftrightarrow p_4$ and $p_2 \leftrightarrow p_3$:

$$I = -\int \mathscr{D}p_1\mathscr{D}p_2\mathscr{D}p_3\mathscr{D}p_4 \left|\mathscr{M}_{\text{f}\to\text{i}}\right|^2 \delta^{(4)}\left(p_1+p_2-p_3-p_4\right)\times\left[n(p_3)n(p_4)-n(p_1)n(p_2)\right]\Phi(p_4). \tag{8.12}$$

Summing these four equations clearly gives

$$\begin{aligned} I &= \int \mathscr{D}p_1\mathscr{D}p_2\mathscr{D}p_3\mathscr{D}p_4 \left|\mathscr{M}_{\text{f}\to\text{i}}\right|^2 \delta^{(4)}\left(p_1+p_2-p_3-p_4\right)\times\left[n(p_3)n(p_4)-n(p_1)n(p_2)\right] \\ &\qquad \times\left[\Phi(p_1)+\Phi(p_2)-\Phi(p_3)-\Phi(p_4)\right]. \end{aligned} \tag{8.13}$$

Reinserting $\Phi(p) = \log n(p)$ and inserting eq. 8.13 into eq. 8.8 then gives

$$\begin{aligned} \frac{ds}{dt} &= -\int \mathscr{D}p_1\mathscr{D}p_2\mathscr{D}p_3\mathscr{D}p_4 \left|\mathscr{M}_{\text{f}\to\text{i}}\right|^2 \delta^{(4)}\left(p_1+p_2-p_3-p_4\right)\times\left[n(p_3)n(p_4)-n(p_1)n(p_2)\right] \\ &\qquad \times\left[\log n(p_1)+\log n(p_2)-\log n(p_3)-\log n(p_4)\right] \\ &= -\int \mathscr{D}p_1\mathscr{D}p_2\mathscr{D}p_3\mathscr{D}p_4 \left|\mathscr{M}_{\text{f}\to\text{i}}\right|^2 \delta^{(4)}\left(p_1+p_2-p_3-p_4\right)\times\left[\underbrace{n(p_3)n(p_4)}_{b}-\underbrace{n(p_1)n(p_2)}_{a}\right] \\ &\qquad \times\log\left(\frac{n(p_1)n(p_2)}{n(p_3)n(p_4)}\right). \end{aligned} \tag{8.14}$$

Boltzmann arrives at his result by recognizing that the dynamical part of this last integral can be written as

$$(b-a)\cdot\log\frac{a}{b}\leq 0 \qquad \Rightarrow \qquad \frac{ds}{dt}\geq 0. \tag{8.15}$$

8.3 Estimate of entropy increase

Now, what kind of entropy increase can we expect? Let's think about the entropy of the universe, that for now we take to be closed (i.e., finite) for simplicity. When the expansion is not too fast, we can try to develop $\frac{dS}{dt}$ in powers of H:

$$\frac{dS}{dt}=\left(c_0+c_1H+c_2H^2+\dots\right)a^3. \tag{8.16}$$

(Here we have used that we expect the entropy to scale proportional to a^3, $S=s\cdot a^3$.) Now, we have to take $c_0=0$, as in thermal equilibrium we have $I_{\text{coll}}=0$, which gives $\frac{dS}{dt}=0$. Also $c_1=0$, because when we go from expansion to contraction, the sign of H changes, and we don't want the entropy to decrease. So we conclude that in first approximation

$$\frac{dS}{dt}=c_2H^2a^3. \tag{8.17}$$

Note that since entropy is dimensionless, the left hand side of this equation has mass dimension one. To have the same dimensionality on the right hand side, it is clear that c_2 will have mass dimension plus two (H has dimension one over time, which means mass dimension one, the scale factor a has dimension length which means mass dimension minus one).

Our next challenge is to estimate c_2, again from using eq. 8.14. We consider the plasma consisting of electrons, positrons and photons[1], and identify three processes:

$$\begin{aligned} e^+e^- &\leftrightarrow \gamma\gamma \\ e^+\gamma &\leftrightarrow e^+\gamma \\ e^-\gamma &\leftrightarrow e^-\gamma. \end{aligned} \tag{8.18}$$

For simplicity, we will again restrict ourselves to the non-relativistic case, even-though the result generalizes to the relativistic case. Introducing the notation $n_+\equiv n_{e^+}(p)$, $n_-\equiv n_{e^-}(p)$ and $n_\gamma\equiv n_\gamma(p)$, we can write three collision integrals for n_+, n_- and n_γ and compute the total entropy density as

$$s_{\text{tot}}=-\int\frac{d^3p}{(2\pi)^3}\left[n_+\log\left(n_+/e\right)+n_-\log\left(n_-/e\right)+n_\gamma\log\left(n_\gamma/e\right)\right]. \tag{8.19}$$

Clearly, the entropy change will be given by eq. 8.14, but with the last factors given by

$$\begin{aligned} &\left[n_\gamma(p_3)n_\gamma(p_4)-n_+(p_1)n_-(p_2)\right]\cdot\log\left[\frac{n_+(p_1)n_-(p_2)}{n_\gamma(p_3)n_\gamma(p_4)}\right] && \text{for} \quad e^+e^-\leftrightarrow\gamma\gamma \\ &\left[n_+(p_3)n_\gamma(p_4)-n_+(p_1)n_\gamma(p_2)\right]\cdot\log\left[\frac{n_+(p_1)n_\gamma(p_2)}{n_+(p_3)n_\gamma(p_4)}\right] && \text{for} \quad e^+\gamma\leftrightarrow e^+\gamma \\ &\left[n_-(p_3)n_\gamma(p_4)-n_-(p_1)n_\gamma(p_2)\right]\cdot\log\left[\frac{n_-(p_1)n_\gamma(p_2)}{n_-(p_3)n_\gamma(p_4)}\right] && \text{for} \quad e^-\gamma\leftrightarrow e^-\gamma. \end{aligned} \tag{8.20}$$

[1] Our approach here is completely general and can handle all types of particles that are present, but for simplicity we focus on the three most abundant particles in the plasma after neutrino decoupling: e^-, e^+ and γ.

Close to thermal equilibrium, we have

$$n \approx n_{\rm eq}\left(1+\mathscr{O}\left(\frac{H}{\Gamma}\right)\right) \qquad \Rightarrow \qquad (n_i n_i - n_i n_i) \simeq n_{\rm eq}^2 \cdot \frac{H}{\Gamma}$$
$$\Rightarrow \qquad \log\left(\frac{n_i\, n_i}{n_i\, n_i}\right) \simeq \frac{H}{\Gamma}. \qquad (8.21)$$

Therefore, eq. 8.14 will give something like (note that now we have the full entropy at the left hand side, $S \propto s \cdot a^3(t)$, with $a(t)$ the (slowly changing) scale factor)

$$\begin{aligned}\frac{dS}{dt} &\simeq \int \mathscr{D}p_1\mathscr{D}p_2\mathscr{D}p_3\mathscr{D}p_4 \,|\mathscr{M}_{\rm f\to i}|^2\, \delta^{(4)}\,(p_1+p_2-p_3-p_4)\times n_{\rm eq}^2\cdot\frac{H^2}{\Gamma^2}\cdot a^3 \\ &\simeq \frac{H^2}{\Gamma} n_{\rm eq}\cdot s\cdot a^3 \\ \frac{1}{S}\frac{dS}{dt} &\simeq \frac{H^2}{\Gamma} n_{\rm eq}. \end{aligned} \qquad (8.22)$$

(Check dimensions: the left hand side goes as one over time, so mass dimension 1. And the same for the right hand side, since both Γ and H have mass dimension 1, while the occupation number is dimensionless. Actually, we have inserted $s \propto T^3$ to get the dimensions right. The qualitative conclusions will not change if instead we use other quantities, like m_e or T, to get the dimensions right.) As anticipated we have found something proportional to H^2.

Now we use that

$$\Gamma = \sigma n v \simeq \sigma n_{\rm eq}\cdot T^3 v \qquad (8.23)$$

(note that the first n is a particle density (mass dimension three), while $n_{\rm eq}$ is (the dominant contribution to) the occupation number (dimensionless), which we integrate over d^3p to get a particle density). We arrive at

$$\frac{1}{S}\frac{dS}{dt} \simeq \frac{H^2}{\sigma T^3}. \qquad (8.24)$$

Now we can compute how much the entropy changes between the moment “in" (decoupling of neutrinos, $T \simeq 2$ MeV) and “out" (freeze out of $e^+e^- \leftrightarrow \gamma\gamma$, $T \simeq \frac{m_e}{40}$, see eq. 6.19 of the lecture notes). Using that we have

$$H \propto \frac{1}{t} \simeq \Gamma = \frac{T^2}{M_0} \qquad (8.25)$$

we can relate time t to temperature T:

$$t = \frac{M_0}{T^2} \qquad \Rightarrow \qquad dt = -\frac{M_0}{T^3}dT, \qquad (8.26)$$

with $M_0 \equiv \frac{m_{\rm Pl}}{1.66 g_*^{1/2}}$. Then we can compute the entropy change as

$$\begin{aligned}\Delta \log S &\propto \int_{t_{\rm in}}^{t_{\rm out}} dt\, \frac{H^2}{\sigma T^3} \\ &\propto -\int_{T_{\rm in}}^{T_{\rm out}} \frac{M_0}{T^3} dT \frac{T^4/M_0^2}{\sigma T^3} \\ &= \frac{1}{M_0}\int_{T_{\rm out}}^{T_{\rm in}} \frac{dT}{T^2\sigma}. \end{aligned} \qquad (8.27)$$

Estimating $\sigma \propto \frac{\alpha^2}{m_e^2}$ gives

$$\begin{aligned}\Delta \log S &\simeq \frac{1}{M_0}\frac{m_e^2}{\alpha^2}\int_{T_{\rm out}}^{T_{\rm in}}\frac{dT}{T^2} \\ &\simeq \frac{1}{M_0}\frac{m_e^2}{\alpha^2}\cdot\frac{1}{T_{\rm out}} \\ &= \frac{1}{M_0}\frac{m_e^2}{\alpha^2}\cdot\frac{40}{m_e} \\ &\approx \frac{m_e}{M_0}\cdot\frac{40}{\alpha^2} \\ &\approx \frac{5\cdot 10^{-4}}{10^{18}}\cdot\frac{40}{10^{-4}} \\ &\approx 10^{-16}. \end{aligned} \tag{8.28}$$

Even if this was a rough estimate, the conclusion is clear: we can, with high accuracy, assume that the entropy (or the entropy density times R^3) is conserved in this period.

8.4 Temperature of Relic Neutrinos

Now let us turn back to physical observables. Not long after the neutrino decoupling, the temperature of the universe goes below electron mass due to the expansion, and e^+'s and e^-'s annihilate into photons. By the time $T \ll m_e$, the remaining particles are γ, ν_e, ν_μ, and ν_τ. The e^+-e^- annihilation process increases the temperature of photons but not that of neutrinos, thus creating a difference between the two temperatures. Let us work out explicitly how much they differ. Suppose that at a time long before the annihilations, when the scale factor is $a_{\rm in}$, both the photons and neutrinos have a temperature $T_{\rm in}$. And, long after the annihilations, when the scale factor is $a_{\rm out}$, the temperature of photons and neutrinos are T_γ and T_ν respectively. Assuming that the universe expands adiabatically, $H \ll \Gamma_W$, we can make use of the conservation of entropy to determine the change in the photon temperature. During the e^+-e^- annihilation, the entropy of e^+'s and e^-'s, which possess $7/8\times(2+2)$ degrees of freedom, is transferred to photons, which possess 2 degrees of freedom. Therefore, the entropy conservation for e^-, e^+, and γ reads

$$\begin{aligned}\left[s_{e^+}+s_{e^-}+s_\gamma\right]\big|_{a=a_{\rm in}} a_{\rm in}^3 &= s_\gamma\big|_{a=a_{\rm out}} a_{\rm out}^3 \\ \left[2+\frac{7}{8}(2+2)\right] T_{\rm in}^3 a_{\rm in}^3 &= 2T_\gamma^3 a_{\rm out}^3\end{aligned} \tag{8.29}$$

Neutrinos, on the other hand, are unaffected by the annihilations, so their temperature simply rescales as

$$T_{\nu,{\rm in}}a_{\rm in} = T_{\nu,{\rm out}}a_{\rm out} \tag{8.30}$$

Combining (8.29) and (8.30), we find the present temperature of relic neutrinos to be

$$T_{\nu,0} = \left(\frac{4}{11}\right)^{1/3} T_{\gamma,0} = 2\ \mathrm{K} \tag{8.31}$$

where $T_{\gamma,0} = 2.73$ K.

Independently of whether the neutrinos are relativistic or non-relativistic today, we can estimate its present number density. The electron mass, and therefore the temperature at which the electrons and positrons annihilated, is much larger than the known upper bounds of the neutrino masses. So, the neutrinos can be assumed to be relativistic close to the time of electron-positron annihilations. We can thus relate the number density of neutrinos with that of photons at a time not long after the annihilations as follows

$$n_\nu = \frac{3}{4}\frac{g_\nu T_\nu^3}{g_\gamma T_\gamma^3} n_\gamma = \frac{3/4}{2}\left(\frac{T_\nu}{T_\gamma}\right)^3 n_\gamma = \frac{3}{22} n_\gamma \tag{8.32}$$

Since both the number density of neutrino and photon scale in the same way as $n \propto a^{-3}$, the above relation is preserved until today.

$$n_{\nu,0} = \frac{3}{22} n_{\gamma,0} \tag{8.33}$$

8.5 Cosmological constraints on neutrino masses

The main constraint on the neutrino masses from cosmology comes from the requirement that the total energy density of the neutrinos is smaller than the observed energy density of dark matter (neutrinos become non-relativistic recently). The energy density of dark matter is

$$\rho_{\rm DM} = \rho_c \Omega_{\rm DM} \tag{8.34}$$

where $\Omega_{\rm DM}$ is the dark matter abundance and for the Hubble constant $H_0 = 100h$ (km/s)/Mpc the critical density ρ_c is

$$\rho_c = \frac{3H_0^2}{8\pi G} = 1.88 \times 10^{-29} h^2 {\rm gcm}^{-3} = 10^4 h^2 {\rm eVcm}^{-3} \tag{8.35}$$

The total energy density of neutrinos today is given by

$$\rho_{\nu,0} = \sum m_\nu n_{\nu,0} = \sum m_\nu \times \left(\frac{3}{22} n_{\gamma,0}\right) \approx \sum m_\nu \times \left(54 \text{ cm}^{-3}\right) \tag{8.36}$$

In the last step, (8.31) and the observed number density of CMB photons $n_\gamma = 400$ cm^{-3} was used. Requiring that $2\rho_{\nu,0} < \rho_{\rm DM}$ (the factor of 2 accounts for antineutrinos), we obtain a bound on the sum of the neutrino masses

$$\sum m_\nu < 100 h^2 \Omega_{\rm DM} \text{ eV} \approx 10 \text{ eV} \tag{8.37}$$

where we have used $h = 0.7$ and $\Omega_{\rm DM} \approx 0.2$. This bound is more stringent than the direct constraints from particle physics:

$$m_{\nu_e} \lesssim 3 \text{ eV}, \quad m_{\nu_\mu} \lesssim 0.17 \text{ MeV}, \quad m_{\nu_\tau} \lesssim 18.2 \text{ MeV} \tag{8.38}$$

Summary:
Shortly after the neutrino decoupling, which occurs at $T \sim 2$ MeV, electrons and positrons in the universe annihilate to produce photons. As a result, the temperature of photons is increased relative to that of neutrinos by a factor of $(11/4)^{1/3}$. The requirement that the total energy density of the neutrinos is less that the observed dark matter energy density gives an upper bound (8.37) on the sum of masses of the neutrinos.

III

Big Bang Nucleosynthesis

Lecture 9

neutron-proton ratio; BBN temperature; ^{4}He abundance;

Roughly speaking, the baryonic matter sector of the current universe consists of

- 75% hydrogen
- 25% helium
- others

in terms of mass. Thermodynamic cooking in stars is far from sufficient to explain the observed amount of Helium and other light elements in the universe, suggesting that a cosmological explanation is needed. Big bang nucleosynthesis (BBN) is our best cosmological explanation for these abundances. With only one parameter as an input, namely the baryon-to-photon ratio η, it explains the observed abundances to high degrees of accuracy.

9.1 Neutron-proton ratio and the BBN temperature

The binding energy B of a nucleus with atomic number A (the number of nucleons comprising the nucleus), charge Z (the number of protons in it), and mass m is given by

$$B = Zm_{\rm p} + (A - Z)m_{\rm n} - m \tag{9.1}$$

See Table 9.1 for examples of binding energy and binding energy per nuclei. The most stable element, one with the highest binding energy per nuclei B/A, is iron (Fe). Given enough time, all nucleons would channel all their resources to form iron. However, it takes an extremely long time for that to happen and consequently most of the nucleons will settle on less stable elements such as deuterium, helium, and lithium.

Due to the difficulties in overcoming potential barrier, heavier elements are produced from lighter ones most effectively through a chain of reactions. Starting from the lightest elements, it goes as follows

$$\begin{aligned}
&{\rm p} + {\rm n} \to {\rm D} + \gamma \\
&{\rm D} + {\rm D} \to {}^3{\rm H} + {\rm p}; \quad {\rm D} + {\rm D} \to^3 {\rm He} + {\rm n} \\
&{}^3{\rm H} + {\rm D} \to {}^4{\rm He} + {\rm n} \\
&{}^3{\rm He} + {\rm n} \to {}^4{\rm H} + {\rm p} \\
&{\rm etc}
\end{aligned} \tag{9.2}$$

Table 9.1: Binding energies and Binding energies per nucleon of some light elements.

nucleus	$B\,(\mathrm{MeV})$	$B/A\,(\mathrm{MeV})$
^{2}H	2.22	1.11
^{3}H	6.92	2.30
^{3}He	7.72	2.57
^{4}He	28.3	7.08
^{12}C	92.2	7.68

In order to describe this chain of reaction, we need to know the starting amount of n and p. These protons and neutrons are those that survived the baryon-antibaryon annihilation occurring in the past, so the sum of their number densities satisfy

$$\eta = \frac{n_B}{n_\gamma} = \frac{n_\mathrm{p} + n_\mathrm{n}}{n_\gamma} = 6.15 \times 10^{-10} \tag{9.3}$$

In fact, the strongest constraint on this number comes from the BBN itself. For the present discussion, all we need to know is that it is a tiny number.

The initial relative abundance of protons and neutrons depends on the details of their freezeout process. The binding energies of light elements are of order 1 MeV, so at $T \gg 1$ MeV but much less than 300 MeV (above which the quarks are free) the particle content of the universe is: p, n, e^+, e^-, γ, ν, and $\bar{\nu}$. The reactions converting protons to neutrons and back are

$$\begin{aligned} &\mathrm{n} \leftrightarrow \mathrm{p} + e^- + \bar{\nu}_e \\ &\mathrm{n} + \nu_e \leftrightarrow \mathrm{p} + e^- \\ &\mathrm{n} + e^+ \leftrightarrow \mathrm{p} + \bar{\nu}_e \end{aligned}$$

Assuming all the species are in thermal equilibrium, we have

$$n_\mathrm{n} \propto \exp\left[-\frac{(m_\mathrm{n} - \mu_\mathrm{n})}{T}\right] \tag{9.4}$$

$$n_\mathrm{p} \propto \exp\left[-\frac{(m_\mathrm{p} - \mu_\mathrm{p})}{T}\right] \tag{9.5}$$

and

$$\mu_\mathrm{n} + \mu_\nu = \mu_\mathrm{p} + \mu_e \tag{9.6}$$

Combining the last three equations, we get

$$\frac{n_\mathrm{n}}{n_\mathrm{p}} = \exp\left(-\frac{Q}{T} + \frac{\mu_e - \mu_\nu}{T}\right) \tag{9.7}$$

with $Q \equiv m_\mathrm{n} - m_\mathrm{p} = 1.293$ MeV. If we further assume that the universe is neutral, $n_e = n_p$, the chemical potential of electron

$$\frac{\mu_e}{T} \approx \frac{n_{e^-} - n_{e^+}}{n_{e^-} + n_{e^+}} \approx \eta \sim 10^{-10} \tag{9.8}$$

can be neglected. The chemical potential of neutrino μ_ν could in principle be large, but we will also neglect it here. Not much is known about the μ_ν, but this assumption is at least justified *a posteriori* by the overwhelming success of BBN in explaining the abundances of light elements in the universe. Since the processes involve only weak interactions, we expect the freezeout temperature of proton and neutron to be roughly the same as what we found earlier in neutrino-decoupling case (8.3), i.e. $T \sim \left(M_0/G_F^2\right)^{1/3} \approx 2$ MeV. More exact calculations gives the proton and neutron freezeout temperature of

$$T_* \approx 0.8 \text{ MeV} \tag{9.9}$$

which corresponds to $t_* \approx 1$ s. Note that t_* is much shorter than the lifetime of neutron $\tau_{\rm n} \approx$ 15 minutes, so its safe to neglect neutrino decay at this stage. After the decoupling, the ratio of proton and neutron abundance essentially freezes at its value at the point of departure from equilibrium

$$\frac{n_{\rm n}}{n_{\rm p}} \approx \left.\frac{n_{\rm n}}{n_{\rm p}}\right|_{T=T_*} = e^{-(m_{\rm n}-m_{\rm p})/T_*} \approx \frac{1}{5} \tag{9.10}$$

This is just the freezeout ratio. As we are going to see now, the ratio will become a different number by the time the BBN starts.

For definiteness, we define the starting point of BBN as the point when the first process in the chain reaction $\rm p + n \leftrightarrow \mathit{D} + \gamma$ becomes effective, that is, when the balance of the reaction starts to shift from the left hand side to the right hand side. We have done a similar calculation when we studied the decoupling of photons. The relevant Saha's equation in the present case is

$$\frac{n_{\rm p} n_{\rm n}}{n_{\rm D}} = \left(\frac{m_{\rm p} T}{2\pi}\right)^{3/2} e^{-\Delta_{\rm D}/T} \tag{9.11}$$

with $\Delta_{\rm D} = m_{\rm p} + m_{\rm n} - m_{\rm D} = 2.23$ MeV. The $\rm p + n \leftrightarrow D + \gamma$ process is said to be effective when $n_{\rm p} \sim n_{\rm D}$. When the latter condition is satisfied, the above Saha's equation reduces to

$$n_{\rm n} \sim \left(\frac{m_{\rm p} T}{2\pi}\right)^{3/2} e^{-\Delta_{\rm D}/T} \tag{9.12}$$

Comparing it with (9.3) and using $n_\gamma \sim T^3$, we find that it occurs when the temperature is

$$T_{\rm BBN} = 70 \text{ keV} \tag{9.13}$$

corresponding to the time $t_{\rm BBN} \approx 4.5$min, which is comparable to the lifetime of neutron. Thus, a significant portion of neutrons must have decayed by then. By the time the deuterium production becomes effective at $T_{\rm BBN}$, the ratio of neutron and proton density has reduced from 1/5 to

$$\frac{n_{\rm n}}{n_{\rm p}} = \exp\left(-\frac{\Delta_{\rm D}}{T_*}\right) \exp\left(-\frac{t_{\rm BBN}}{\tau_{\rm n}}\right) \approx \frac{1}{7} \tag{9.14}$$

Soon after that, most of the neutrons in the universe will be hidden inside deuteriums, preventing them from decaying any further.

9.2 BBN's prediction for the ^{4}He abundance

The abundance of nuclei with atomic number A is often presented in terms of the quantity

$$x_A = \frac{A n_A}{n_B} \tag{9.15}$$

where n_A is the number density of the nuclei in question and n_B is the total number density of baryons. Throughout this section, when we say the abundance of an element we mean exactly this number. By definition, the abundances of all nuclei should sum to 1, $\sum_A x_A = 1$. The ^{4}He nucleon is special because its binding energy per nucleon $B_A/A = 7$ MeV happens to be a local maximum among the neighboring nuclei in the (A,Z) space. Consequently, a large part of the protons and neutrons end up forming ^{4}He (see the remark at the end of this section for a more rigorous justification). Considering that a ^{4}He nucleon consists of 2 protons and 2 neutrons and that there are less neutrons than protons, the formation of ^{4}He will stop once all the neutrons have been exhausted. Knowing this, we can estimate the amount of ^{4}He as follows

$$n_{^4\mathrm{He}} = \frac{1}{2} n_n \tag{9.16}$$

The corresponding abundance can be computed with the aid of the neutron to proton ratio (9.14) we obtained earlier

$$x_4 = \frac{4n_{^4\mathrm{He}}}{n_p + n_n} = \frac{2n_n/n_p}{1+n_n/n_p} = \frac{2\times 1/7}{1+1/7} = \frac{1}{4} = 25\% \tag{9.17}$$

which accurately explains the observed value if the calculation is done carefully. This prediction is robust in that it does not depend on the baryon to photon ratio η. The abundances of other less abundant elements, e.g. deuterium, lithium, etc, hinge solely on η and are well explained if η is given by (9.3), thus providing one of the strongest evidences for the big bang theory.

Remark: more rigorous reason why essentially all the neutrons end up forming $n_{^4\mathrm{He}}$.
We start by writing down the equilibrium number densities of the species we want to consider

$$\begin{aligned}
n_\mathrm{p} &\propto \exp\left[-\frac{(m_\mathrm{p}-\mu_\mathrm{p})}{T}\right] \\
n_\mathrm{n} &\propto \exp\left[-\frac{(m_\mathrm{n}-\mu_\mathrm{n})}{T}\right] \\
n_\mathrm{D} &\propto \exp\left[-\frac{(m_\mathrm{D}-\mu_\mathrm{D})}{T}\right] \\
n_\mathrm{T} &\propto \exp\left[-\frac{(m_\mathrm{T}-\mu_\mathrm{T})}{T}\right] \\
n_{^3\mathrm{He}} &\propto \exp\left[-\frac{(m_{^3\mathrm{He}}-\mu_{^3\mathrm{He}})}{T}\right] \\
n_{^4\mathrm{He}} &\propto \exp\left[-\frac{(m_{^4\mathrm{He}}-\mu_{^4\mathrm{He}})}{T}\right]
\end{aligned}$$

etc

The more species we include, the better. Then, list reactions keeping them in thermal equilibrium and the corresponding chemical equilibrium conditions

$$\begin{aligned}
&\mathrm{p}+\mathrm{n} \leftrightarrow \mathrm{D}+\gamma, && \mu_\mathrm{p}+\mu_\mathrm{n} = \mu_\mathrm{D} \\
&^3\mathrm{H}+\mathrm{p} \leftrightarrow \mathrm{D}+\mathrm{D}, && \mu_\mathrm{T}+\mu_\mathrm{p} = \mu_\mathrm{D}+\mu_\mathrm{D} \\
&^3\mathrm{He}+\mathrm{p} \leftrightarrow \mathrm{D}+\mathrm{D}, && \mu_{^3\mathrm{He}}+\mu_\mathrm{p} = \mu_\mathrm{D}+\mu_\mathrm{D} \\
&^3\mathrm{H}+\mathrm{D} \leftrightarrow ^4\mathrm{He}+\mathrm{n}, && \mu_T+\mu_\mathrm{D} = \mu_{^4\mathrm{He}}+\mu_\mathrm{n}
\end{aligned}$$

etc

For an arbitrary element A, we have the following relationship

$$\mu_A = \mu_\mathrm{p} \times (\text{num. of protons}) + \mu_\mathrm{n} \times (\text{num. of neutrons}) \tag{9.18}$$

Thus, the number density of the element is given by

$$n_A = g_A \left(\frac{m_A T}{2\pi}\right)^{3/2} \exp\left[-\frac{Z_A(m_\mathrm{p}-\mu_\mathrm{p})}{T} - (A_A - Z_A)\frac{(m_\mathrm{n}-\mu_\mathrm{n})}{T} + \frac{B_A}{T}\right] \tag{9.19}$$

Knowing the total numbers of protons and neutrons, we impose the following conditions

$$\sum n_A Z_A = \frac{n_\mathrm{p}}{n_\mathrm{n}+n_\mathrm{p}}\eta n_\gamma \approx \frac{6}{7}n_\gamma \tag{9.20}$$

$$\sum n_A (A_A - Z_A) = \frac{n_\mathrm{n}}{n_\mathrm{n}+n_\mathrm{p}}\eta n_\gamma \approx \frac{1}{7}n_\gamma \tag{9.21}$$

Combining (9.19) for protons and (9.20) gives

$$\begin{aligned}\left(\frac{m_\mathrm{p} T}{2\pi}\right)^{3/2} \exp\left[-\frac{(m_\mathrm{p}-\mu_\mathrm{p})}{T}\right] &\sim \eta n_\gamma \\ \exp\left[-\frac{(m_\mathrm{p}-\mu_\mathrm{p})}{T}\right] &\sim \eta \left(\frac{T}{m_\mathrm{p}}\right)^{3/2}\end{aligned} \tag{9.22}$$

Similarly, combining (9.19) for ^{4}He and (9.21) gives

$$\begin{aligned}\left(\frac{m_\mathrm{p} T}{2\pi}\right)^{3/2} \exp\left[-2\frac{(m_\mathrm{p}-\mu_\mathrm{p})}{T}\right] \exp\left[-2\frac{(m_\mathrm{n}-\mu_\mathrm{n})}{T}\right] \exp\left(\frac{B_{^4\mathrm{He}}}{T}\right) &\sim \eta n_\gamma \\ \exp\left[-\frac{(m_\mathrm{n}-\mu_\mathrm{n})}{T}\right] &\sim \left(\frac{m_\mathrm{p}}{T}\right)^{3/4} \frac{1}{\sqrt{\eta}} \exp\left(-\frac{B_{^4\mathrm{He}}}{2T}\right)\end{aligned} \tag{9.23}$$

At $T \approx 100$ keV, we have

$$\exp\left[-\frac{(m_\mathrm{n}-\mu_\mathrm{n})}{T}\right] \sim 10^{-53} \tag{9.24}$$

which is extremely small and typically gives more dominant effect than the factor (9.22). Having (9.22) and (9.23) in hand, we can estimate the number density of any element. For instance, the number density of deuterium relative to ^{4}He is

$$\frac{n_\mathrm{D}}{n_{^4\mathrm{He}}} \sim \exp\left(\frac{m_\mathrm{p}-\mu_\mathrm{p}}{T} + \frac{m_\mathrm{n}-\mu_\mathrm{n}}{T} + \frac{B_\mathrm{D}}{T} - \frac{B_{^4\mathrm{He}}}{T}\right) \sim \exp(-118) \ll 1 \tag{9.25}$$

which is essentially zero. Doing the same to all the other elements will lead us to the conclusion that pretty much all the neutrons end up in ^{4}He.

Summary:
BBN is our best cosmological explanation for the observed abundance of light elements in the universe for which thermodynamic cooking in stars does not account. Before their decoupling, the protons and neutrons that survived baryon-antibaryon annihilation were held in thermal equilibrium by the process $\mathrm{p}^+ + e^- \leftrightarrow \mathrm{n} + \nu_e$ and its variants. After the decoupling, the neutron to proton ratio essentially freezes at its value at the departure from equilibrium, $n_n/n_p \approx 1/5$. Once the temperature of the universe reduces to below the binding energy of helium, it becomes thermodynamically favourable for the helium nucleus sector to eat up all the protons and neutrons they can. While the lifetime of proton is longer than the age of the universe, significant amount of neutrons decayed in the mean time, changing the neutron to proton ratio to $n_n/n_p \approx 1/7$. After the eating up, the total mass of helium makes up 25% of the total mass of baryons, in agreement with observations.

IV

Baryogenesis

Lecture 10

baryon to entropy ratio η; Sakharov's conditions; GUT baryogenesis

10.1 Baryon to entropy ratio

It is an observed fact that the present universe has overwhelmingly more baryons than antibaryons. Rather than taking this fact for granted, most physicists find it more natural to assume that the universe started in a baryon-symmetric state and somehow produced baryon asymmetry as it evolved. This process of creating baryon asymmetry is dubbed baryogenesis. Once the temperature goes below the electroweak scale, ~ 100 GeV, the baryon number is known to be conserved to a very good approximation. In an expanding universe, this translates to

$$(n_B - n_{\bar{B}})a^3 = \text{const} \tag{10.1}$$

Furthermore, we found before that as long as the universe expands adiabatically, the entropy is conserved

$$sa^3 = \text{const} \tag{10.2}$$

If we take the ratio between (10.1) and (10.2), the scale factors would cancel and we get

$$\eta \equiv \frac{n_B - n_{\bar{B}}}{s} = \text{const} \tag{10.3}$$

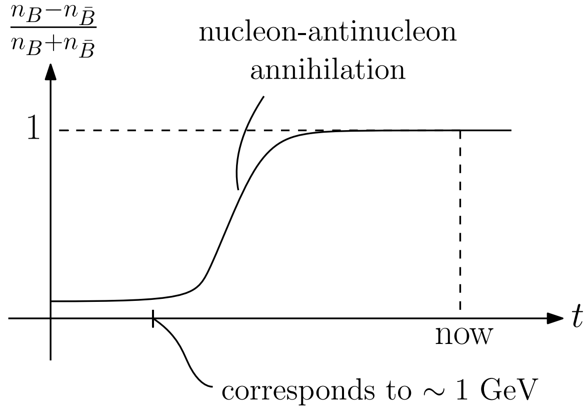
Thanks to this fact, while the process of baryogenesis is not well understood one could skip its details and take the produced baryon to entropy ratio η as an input parameter for the subsequent evolutions. This was exactly what we have done when we studied BBN. In this lecture, we will attempt to understand the origin of the baryon asymmetry in our universe.

At temperatures larger than the masses of nucleons, i.e. $T \gtrsim 1$ GeV, photons, baryons, and antibaryons were all thermally populated with number densities $n_B \sim n_{\bar{B}} \sim n_\gamma \sim T^3$. At that time, the baryon to entropy ratio is given by

$$\eta \equiv \frac{n_B - n_{\bar{B}}}{s} \sim \frac{n_B - n_{\bar{B}}}{T^3} \sim \left.\frac{n_B - n_{\bar{B}}}{n_B + n_{\bar{B}}}\right|_{T \gtrsim 1\ \text{GeV}} \tag{10.4}$$

In this sense η gives a measure of the baryon asymmetry at $T \gtrsim 1$ GeV. Then as the temperature cools down pass 1 GeV, the baryons and antibaryons annihilate to produce photons. They, of course,

Figure 10.1: Schematic behaviour of $(n_B - n_{\bar{B}})/(n_B + n_{\bar{B}})$ as function of time.



did not annihilate completely. The observed baryon to entropy ratio of $\eta \sim 10^{-10}$ indicates that for every 10^{10} baryons and antibaryons, there were about one baryon which could not annihilate, resulting in about 10^{10} more photons than baryons today (see Figure 10.1). Since the current temperature of the universe $T_0 \sim 1$ meV is very low, the entropy density is accurately given by the contribution from photons, the only relativistic particles at the time

$$s \approx \frac{\zeta(3)}{\pi^2} g_\gamma T^3 \sim n_\gamma \tag{10.5}$$

This, together with the fact that we have $n_B \gg n_{\bar{B}}$ today, allows us to write

$$\eta \sim \frac{n_B}{n_\gamma} \tag{10.6}$$

10.2 Sakharov's conditions

Reminder: P, T, and C.

- The parity transformation P is a spacetime transformation that inverts space $\mathbf{x} \to -\mathbf{x}$ while leaving time unchanged $t \to t$. Different physical variables transform differently under P, e.g. the momentum $\mathbf{p}$ transforms as $\mathbf{p} \to -\mathbf{p}$ and the angular momentum $\mathbf{M}$ transforms as $\mathbf{M} \to \mathbf{M}$.
- The time-reversal transformation T is a spacetime transformation that inverts time $t \to -t$ while leaving space unchanged $\mathbf{x} \to \mathbf{x}$.
- The charge conjugation C turns particles into antiparticles and vice versa.

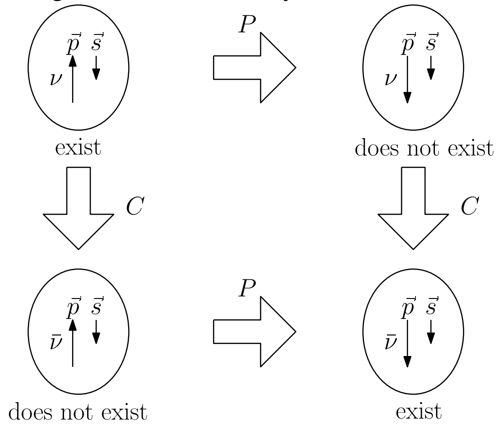
People used to believe that C and P were exact symmetries until the 1956 discovery of C and P non-conservation in decay processes involving a neutrino showed otherwise. If C and P were respected, all four different processes shown in Figure 10.2 with different directions of momentum and spin of the neutrino decay product would occur equally likely. It turns out that only two of them could happen. Thus, in this case both C and P are maximally violated. After that, physicists no longer regard C and P as fundamental symmetries, but still expected the combination CP to be exact. In 1964, CP-violation in decays involving K_L^0 mesons were observed, shocking the community once more. Among the possible ways K_L^0 could decay are

$$K_L^0 \to \pi^- + e^+ + \nu_e \tag{10.7}$$

$$K_L^0 \to \pi^+ + e^- + \bar{\nu}_e \tag{10.8}$$

CP transformation converts the decay products of the first process into those of the second one. Thus, if we prepare equal amounts of K_L^0 and $\bar{K}_L^0$, we expect to see equal amounts of, say, e^-

Figure 10.2: Discovery of C and P violation.



and e^+ in the end. It was found that

$$\frac{n_{e^+} - n_{e^-}}{n_{e^+} + n_{e^-}} \sim O(10^{-3}) \tag{10.9}$$

which implies that CP is violated. Now that CP is known to be broken, we are still hopeful that CPT is an exact symmetry. CPT conservation implies that the mass and lifetime of a particle must be the same as those of the antiparticle.

Remark: a universal way to distinguish particles from antiparticles.
The choice of labeling a particle pair which are related to each other via CP transformation as matter and antimatter is obvious when one member of the pair is much more abundant than the other, in which case we simply name the more abundant one matter and the less abundant one antimatter. Now, the question is how do we single out what we have been referring to as matter from antimatter somewhere in the universe where they are equally abundant? One way to do so is to make use of the CP-violating decays of K_L^0: prepare an equal amount of K_L^0 and $\bar{K}_L^0$, and let them decay; call positron a charged lepton which appears more often in the decay and call proton a baryon with the same sign of charge.

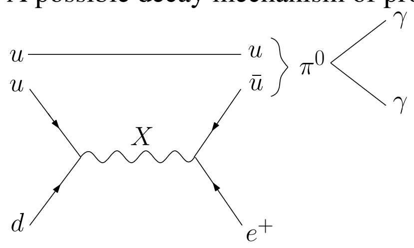
From the viewpoint of particle physics, in order to produce a net baryon asymmetry in the universe the three so-called Sakharov conditions must be met:

1. Existence of baryon number violating process.
   Although it seems inconsistent with the stability of matter, e.g. the lifetime of proton was found to be longer than 10^{31} years, baryon number violation might occur at high energies.
2. Departure from thermal equilibrium.
   Otherwise, we would have $n_B(p) = n_{\bar{B}}(p) = \frac{1}{e^{E(p)/T} - 1}$, and no baryon asymmetry would be created. We did not include any chemical potential because in this context the baryon number is assumed to be not conserved.
3. Violation of C and CP.
   This is simply because while baryon-symmetric states are C and CP invariant, baryon-asymmetric states break C and CP. So, in order to go from a baryon-symmetric state to a baryon-asymmetric one, the laws of physics have to break C and CP.

10.3 GUT baryogenesis

Reminder: a particle physics interlude.

Figure 10.3: A possible decay mechanism of proton in GUT.



A way to specify a theory in terms of its Lagrangian, which is required to be invariant under every symmetry of the theory. To write down the Standard Model, we begin by choosing the symmetry group. In addition to the Poincare symmetry, the Standard Model respects the gauge symmetries

$$G = SU(3) \times SU(2) \times U(1)$$

Once we specified the gauge group, the gauge bosons in the theory are fixed. For each generator of the group there is a gauge boson that transforms according to the adjoint representation of the group. We have one photon, three W bosons, and eight gluons gauge bosons corresponding to the $U(1)$, $SU(2)$, and $SU(3)$ symmetries respectively. Next, we choose the matter content of our model, together with how they interact, i.e. the way they transform under the gauge group G. Particles are representations of the Poincare group which are classified by their spins (scalars, fermions, gauge bosons) and the way they transform under the symmetries (the charges). For some unknown reason, each of the fermions of the theory comes in 3 copies (generations), which are identical to one another apart from their masses. For example, the fundamental building blocks of baryons are quarks, which come in pairs: (u,d), (c,s), and (t,b). The "up" quarks (u, c, t) each carries an electric charge of $q = 2/3$ and a baryon number of $b = 1/3$, while the "down" quarks (d, s, b) each carries an electric charge $q = -1/3$ and a baryon number $b = 1/3$. Each nucleons consists of three quarks $p = uud$, $n = udd$. Mesons are constructed out of a quark and an antiquark $q\bar{q}$. Due to the so-called quark confinement, quarks do not exist freely in ordinary conditions (small temperature and density).

The economy of description or simplicity has always been a quality that physicists sought after in their theories. Looking at the group theoretic structure of the Standard Model, one is naturally led into questioning whether the electromagnetic, weak, and strong interactions would merge into one at some high energy scale. Grand unified theories (GUTs) are models that realize such a unification. As a bonus, GUTs are also attractive because they often automatically provide possible baryogenesis mechanisms. In one of the simplest GUT models based on the $SU(5)$ symmetry, there are particles called leptoquarks, denoted by X, which can undergo baryon-number violating processes into quarks and leptons. For example, X could decay through the following processes: $X \to u + d$ and $X \to e^+ + \bar{u}$. As a consequence, the model predicts that proton is unstable. It could, for instance, undergo the process shown in Figure 10.3, with a decay width

$$\Gamma \sim \frac{g^4}{M_X^4} m_p^5 \sim \frac{\alpha^2 m_p^5}{M_x^4} \tag{10.10}$$

The current experimental constraint for the lifetime of proton is $\tau_p \gtrsim 10^{31}$ years, meaning that the leptoquarks X must be very heavy if they exist, i.e. $M_X > 10^{15}$ GeV.

A possible baryogenesis scenario based on the $SU(5)$ GUT goes as follows. At $T \gtrsim 10^{15}$ GeV, both leptoquarks and antileptoquarks were presumably in thermal equilibrium, having equal abundances. The idea is that their decay would slightly favor particles over antiparticles, thus creating baryon asymmetry in the universe. There are four types of (CP-violating) leptoquark decay: $X \to qq$, $X \to \bar{q}\bar{\ell}$, $\bar{X} \to \bar{q}\bar{q}$, and $\bar{X} \to q\ell$. While CP-violation allows

$$\Gamma(X \to qq) \neq \Gamma(\bar{X} \to \bar{q}\bar{q}) \tag{10.11}$$

The CPT theorem demands

$$\Gamma(X \to qq) + \Gamma(X \to \bar{q}\bar{\ell}) = \Gamma(\bar{X} \to \bar{q}\bar{q}) + \Gamma(\bar{X} \to q\ell) \tag{10.12}$$

or

$$\Gamma(X \to qq) - \Gamma(\bar{X} \to \bar{q}\bar{q}) = \Gamma(\bar{X} \to q\ell) - \Gamma(X \to \bar{q}\bar{\ell}) \tag{10.13}$$

Now, suppose that we have one X and one $\bar{X}$. Let us see how much baryon number asymmetry will be created after their decays. On one hand, the decay of X produces

$$B_X = \frac{2}{3}\frac{\Gamma(X \to qq)}{\Gamma_{\text{tot}}} - \frac{1}{3}\frac{\Gamma(X \to \bar{q}\bar{\ell})}{\Gamma_{\text{tot}}} \tag{10.14}$$

The factor $2/3$ and $-1/3$ in the first and second terms are there because two quarks are produced in the $X \to qq$ process and one antiquark is produced in the $X \to \bar{q}\bar{\ell}$ process. On the other hand, the decay of $\bar{X}$ produces

$$B_{\bar{X}} = -\frac{2}{3}\frac{\Gamma(\bar{X} \to \bar{q}\bar{q})}{\Gamma_{\text{tot}}} + \frac{1}{3}\frac{\Gamma(\bar{X} \to \bar{q}\bar{\ell})}{\Gamma_{\text{tot}}} \tag{10.15}$$

So, overall, we have

$$B_{\text{tot}} = B_X - B_{\bar{X}} = \frac{\Gamma(X \to qq) - \Gamma(\bar{X} \to \bar{q}\bar{q})}{\Gamma_{\text{tot}}} \tag{10.16}$$

If there were n_X leptoquarks and antileptoquarks, we would have

$$\frac{n_B}{n_\gamma} \approx B_{\text{tot}}\frac{n_X}{n_\gamma} \tag{10.17}$$

10.4 Particle decays in GUT baryogenesis

We found in the previous section that if the baryon asymmetry in the universe was originated from the decay of leptoquarks, then it would depend on the starting amount of the leptoquarks n_X, more specifically on the ratio n_X/s, at the moment of decay. Assuming the leptoquarks were in thermal equilibrium, we have

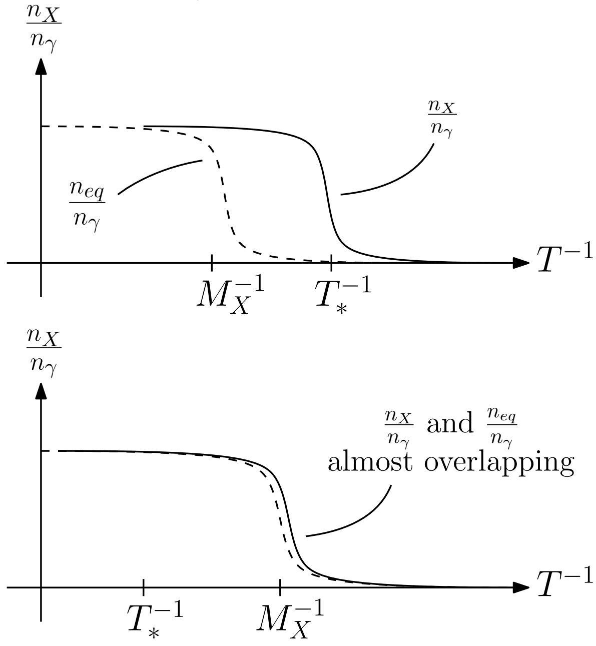
$$\frac{n_X}{n_\gamma} \sim \begin{cases} 1, & T \gg M_X \\ \left(\frac{M_X}{T}\right)^{3/2} \exp\left(-\frac{M_X}{T}\right), & T \ll M_X \end{cases} \tag{10.18}$$

If the leptoquarks continues to be in thermal equilibrium, the processes $X \to qq$ and $X \to \bar{q}\bar{\ell}$ and their reverse processes $qq \to X$ and $\bar{q}\bar{\ell} \to X$ would remain equally likely. Thus, in order for the GUT baryogenesis to work the leptoquarks must have decoupled from the rest of the thermal bath before they start to decay, which is basically one of the Shakarov's conditions. The decay rate of leptoquarks is roughly given by

$$\Gamma_{\text{tot}} \sim \begin{cases} \Gamma_{\text{tot}}\frac{M_X}{T}, & T \gg M_X \\ \Gamma_{\text{tot}}, & T \lesssim M_X \end{cases} \tag{10.19}$$

The factor $M_X/T \sim 1/\gamma$ is a Lorentz factor that captures the fact that, due to time dilation, relativistic particles decay at significantly slower rates. A priori, the leptoquarks could either decay when they are relativistic or when they are non-relativistic. Since M_X is a very weakly constrained quantity, we should consider both possibilities.

Figure 10.4: Rough behavior of n_X/n_γ in the $T_* \lesssim M_X$ case (top) and $T_* \gg M_X$ case (bottom).



Comparing (10.19) with the now familiar $H = T^2/M_0$, we find that the decay happens at

$$T_* \sim \begin{cases} (\Gamma_{\text{tot}} M_X M_0)^{1/3}, & T_* \gg M_X \\ \sqrt{\Gamma_{\text{tot}} M_0}, & T_* \lesssim M_X \end{cases} \tag{10.20}$$

Thus, at T_* the leptoquarks are relativistic ($T_* \gg M_X$) if

$$\frac{M_X}{\Gamma_{\text{tot}} M_0} \gg 1 \tag{10.21}$$

and non-relativistic ($T_* \lesssim M_X$) if

$$\frac{M_X}{\Gamma_{\text{tot}} M_0} \lesssim 1 \tag{10.22}$$

Let us first explore the non-relativistic $T_* \lesssim M_X$ case. Figure 10.4 shows the behavior of n_X/n_γ in this case. As we can see in the figure, the thermal equilibrium ratio $n_{X,\text{eq}}$ starts to be Boltzmann suppressed at $T \sim M_X$, but at that time the leptoquarks cannot decay fast enough to track the thermal equilibrium value. Consequently, n_X/n_γ maintains its initial value ~ 1, thus departing from the now Boltzmann-suppressed equilibrium value, until the temperature drops down to $T \sim T_*$, at which point significant amount of leptoquarks start to decay within a Hubble time. At the moment of the decay $n_X/n_\gamma \sim 1$ and so by (10.17) we find the baryon to photon ratio created in the decay process to be

$$\frac{n_B}{n_\gamma} \sim \frac{\Gamma(X \to qq) - \Gamma(\bar{X} \to \bar{q}\bar{q})}{\Gamma_{\text{tot}}} \frac{1}{g_*} \tag{10.23}$$

where g_* is the effective number of relativistic degrees of freedom. If we count all quarks, leptons and gauge bosons, $g_* \sim 100$.

We now turn to the relativistic $T_* \gg M_X$ case. As shown in Figure 10.4, n_X/n_γ has no trouble following its thermal equilibrium abundance closely with $n_X/n_\gamma \approx n_{X,\text{eq}}/n_\gamma$ until $T \sim T_*$. So, at the moment of decay, one of the Sakharov conditions, namely departure from thermal equilibrium, is not fulfilled. We are therefore led to conclude that no baryon asymmetry is produced in the decay process. Strictly speaking, due to the finite time it takes to achieve thermal equilibrium n_X/n_γ is bound to slightly lag behind its thermal equilibrium value. The deviation from thermal equilibrium is roughly given by

$$\frac{n_X - n_{X,\mathrm{eq}}}{n_\gamma} \sim O\left(\frac{H}{\Gamma_{\mathrm{tot}}}\right) = \frac{T^2/M_0}{\Gamma_{\mathrm{tot}}} \sim \frac{M_X^2}{M_0\Gamma_{\mathrm{tot}}} \tag{10.24}$$

In the last step, we have set $T \sim M_X$, which is roughly the point where $n_{X,\mathrm{eq}}/n_\gamma$ starts to drop rapidly due to Boltzmann suppression. The amount of baryon asymmetry created in this case will be suppressed by a factor of $M_X^2/M_0\Gamma_{\mathrm{tot}} \ll 1$ relative to what we found in the $T_* \lesssim M_X$ case. So, in conclusion significant baryon asymmetry could only be created if the leptoquarks decay occurs when they are non-relativistic, i.e. $T_* \lesssim M_X$. The constraint on M_X coming from this requirement can be calculated by imposing (10.22) and taking

$$\Gamma_{\mathrm{tot}} \sim \alpha M_X \tag{10.25}$$

with $\alpha \sim 10^{-2}$. These give

$$M_X \gtrsim \alpha M_0 \sim 10^{17}\ \mathrm{GeV} \tag{10.26}$$

which is incidentally quite close and compatible to the current bound coming from proton the decay experiments.

Remark: some comments on baryogenesis in general.
Qualitatively speaking, the Standard Model has all the necessarily ingredients to create baryon asymmetry: the CKM matrix that describes the mixing among the different quark flavours contains a CP-violating phase, departure from thermal equilibrium can be achieved naturally in an expanding universe, and baryon-number violation may occur through the so-called sphaleron process. The sphaleron process is a non-perturbative process that does the following conversion

$$9q + 3\ell \to \text{bosons}$$

The process takes place when there is a transition from one minimum to another. At $T = 0$, the transition could happen through quantum tunneling, but with a highly suppressed probability $\sim \exp(-4\pi/\alpha)$. However, if the temperature is finite $T \neq 0$ and sufficiently high, the transition if feasible classically by climbing over the barrier with the help thermal fluctuations with probability $\sim \exp(-M_{\mathrm{sph}}/T)$, where $M_{\mathrm{sph}} \sim M_W/\alpha$ is the mass of sphaleron. The probability becomes significant when $T \gtrsim M_{\mathrm{sph}} \sim 150$ GeV. These are all fine qualitatively, but when we actually calculate the amount of baryon asymmetry created through the said process, the number is far too small to account for the asymmetry that we observed.

The conclusion we draw from the preceding discussion is that, in order to explain baryogenesis, new and stronger sources of CP violations coming from beyond the Standard Model (BSM) physics are needed. One possibility is by extending the Standard Model with three right-handed neutrinos. Since these additional neutrinos are very weakly coupled to the Standard Model sector, they are naturally out of equilibrium. Their decay produce lepton asymmetry, which could then be converted to baryon asymmetry via sphaleron processes.

Another possibility is to extend the scalar and fermionic sectors of the Standard Model in such a way that first order electroweak phase transition could occur. Suppose that an electroweak phase transition took place at some point in the cosmological history. When the transition took place, it would not occur simultaneously everywhere in the universe. In the beginning of the transition, the regions where the phase transition had taken place showed up as bubbles. These bubbles would then grow in size and eventually percolate. When the bubbles were growing in size, the surfaces (domain walls) of the bubbles sweep over quarks and antiquarks they come in contact with. The bubble surface lets leptons enter more easily than antileptons. So, in the end we would have a universe with excess leptons over antileptons. This lepton excess could then be converted to baryon excess that we see today through sphaleron processes. Unfortunately, with the recent measurement of the Higgs mass we found out that the electroweak phase transition should not have happened, unless we are still missing something in the Standard Model that significantly changes the electroweak-sector phase diagram.

Summary:
The universe as we observe today appears to be populated exclusively by matter rather than antimatter. This poses a puzzle if we start with the assumption that the early universe has equal amounts of matter and antimatter. To produce a net baryon asymmetry we require three conditions (Sakharov conditions): the existence of baryon violating process, C and CP violation, and departure from thermal equilibrium. The first two are expected to happen at high energies and the third one is achievable in an expanding universe. There are many viable mechanisms of baryogenesis, but experiments are yet to single them out. One possible baryogenesis mechanism is through the decay of leptoquarks in the $SU(5)$ Grand Unified Theory (GUT).

V

Dark Matter

Lecture 11

> evidences for dark matter; dark matter candidates within the Standard Model; beyond the Standard Model dark matter candidates; WIMP dark matter; known properties of dark matter; unknown properties of dark matter.

11.1 Evidences for dark matter

The existence of non-luminous and non-absorbing substance, dubbed dark matter, making up about 27% of the mass of the universe is now an established fact. Our confidence on its existence has been built on many independent observations pointing to the same conclusion. Here we discuss some of them.

11.1.1 Rotational curves of spiral galaxies

The first compelling evidence for dark matter concerns the orbital velocities of stars located on the disk of a spiral galaxy, see Figure 1.1 for a rough sketch of a spiral galaxy. The mass distribution of luminous matter in such a galaxy can be inferred by measuring its luminosity as function of radius. Observations show that the luminosity follows a rough radial dependence of the form

$$I(r) = I_0 \exp\left(-\frac{r}{r_0}\right) \tag{11.1}$$

Based on that, we can infer that most of the mass in a spiral galaxy is concentrated near its center. If we assume that only luminous matter gravitate, we expect the velocities of the stars on the disk at the outer side of the galaxy to fall off according to the Kepler's law ($v \propto 1/\sqrt{r}$). The reason is that, since most of the mass is clumped near the center of the galaxy, stars located at different radii sufficiently far from the center of the galaxy feel essentially the same amount of mass pulling them towards the center. Mathematically,

$$\begin{aligned} mv^2 &\approx G\frac{Mm}{r} \\ v &\approx \sqrt{\frac{GM}{r}} \end{aligned} \tag{11.2}$$

In reality, what we see is a flat profile

$$v \approx \text{constant} \tag{11.3}$$

suggesting that the M in (11.2) is not a constant but varies as

$$M(r) \propto r \tag{11.4}$$

which could be explained by the presence of a dark matter halo with mass distribution

$$\rho_{\text{dark}}(r) \propto \frac{1}{r^2} \tag{11.5}$$

Based on observations, the dark matter density close to the center of a spiral galaxy is roughly uniform. A better form of $\rho_{\text{DM}}(r)$ that captures it qualitative behaviours at $r \to 0$ and $r \to \infty$ is

$$\rho_{\text{DM}}(r) \approx \frac{v_\infty^2}{4\pi G r_c^2} \frac{r_c^2}{r^2 + r_c^2} \tag{11.6}$$

11.1.2 Gravitational lensing

Gravitational lensing is a phenomenon where massive objects located between us and distant light sources act as a lens, bending the space around them and consequently bending any light passing nearby. When the effect only causes slight shear deformations in the images of distant luminous objects, it is called weak gravitational lensing. By measuring the amount of such shear deformations on distant galaxies, and combining it with appropriate statistical analysis, we can infer the total mass of the intervening galaxy clusters causing the deformations. This led to the same conclusion that the galaxy clusters are more massive than the total mass of the visible objects belonging to them.

11.1.3 Bullet Cluster

The Bullet Cluster is an aftermath of a collision between two galaxy clusters. The mass distribution of the colliding clusters can be inferred by gravitational lensing. It was found that the non-luminous matter simply pass through one another, showing that the dark matter has essentially no pressure. Furthermore, it was observed that the location of the center of mass inferred from gravitational lensing is significantly displaced from that of the (luminous) baryonic matter, a fact that can be explained easily by the presence of dark matter but not by modifying the theory of gravity, thus favouring the former over the latter. In particular, the Bullet Cluster observation ruled out the simplest version of MOND.

11.1.4 Big Bang nucleosynthesis

The BBN scenario predicts the abundances of light elements in the universe depending on the abundance Ω_B of baryons in the universe. Requiring the predictions to match with the observed values put a stringent bound on the value of Ω_B. It was found that Ω_B is smaller than the observed abundance Ω_M of non-relativistic matter in the universe, which suggests the existence of dark matter.

11.1.5 Combined data from SNe, BAO, and CMB

The combined data from supernovae (SNe), baryon acoustic oscillation (BAO), and the cosmic microwave background (CMB) constraint the abundances of non-relativistic matter Ω_M, non-relativistic matter in the form of baryons Ω_B, and the cosmological constant Ω_Λ to have the following values

$$\Omega_M \approx 0.317, \quad \Omega_B \approx 0.049, \quad \Omega_\Lambda \approx 0.683 \tag{11.7}$$

The fact that $\Omega_B < \Omega_M$ by a significant amount suggest the presence of non-baryonic matter, i.e. dark matter.

11.2 Standard Model (failed) dark matter candidates

The particle nature of the dark matter is not yet known. For the reasons that we will provide below, we can at least say that, most probably, the dark matter is not made of Standard Model particles.

The only electrically-neutral and stable objects in the Standard Model are atoms and neutrinos. The dark matter cannot be in the form of atoms (baryons) because the abundance Ω_B of baryons in the universe has been pinned down by BBN considerations to around one fifth of the observed dark matter abundance, $\Omega_B \approx \Omega_{\rm DM}/5$; if Ω_B were to increase only slightly, the BBN predictions on the abundances of light elements in the universe would differ significantly from the observed values.

The dark matter cannot be the Standard Model neutrinos for the following reason. The Pauli exclusion principle gives an upper limit on how densely neutrinos can be packed in the phase space $(\mathbf{p},\mathbf{x})$. Consequently, the number of neutrinos N_ν with typical velocity v contained within an object of size r, e.g. a galaxy halo, is bounded from above

$$N_\nu \lesssim \frac{1}{(2\pi)^3}\int d^3\mathbf{p}d^3\mathbf{x}n \sim p^3 r^3 \tag{11.8}$$

In the last step we have assumed that n has the form of a window function with momentum and coordinate extent of p and r respectively. It follows that the total mass of neutrinos in a galaxy M is also bounded from above

$$M \lesssim m_\nu N_\nu \sim m_\nu^4 v^3 r^3 \tag{11.9}$$

At the same time, the Kepler's law states that $v^2 \sim GM/r$, which, when combined with the above bound, gives an upper limit on the masses of neutrinos in terms of measurable quantities v and r

$$m_\nu \gtrsim \left(\frac{1}{Gvr^2}\right)^{1/4} \approx 120\left(\frac{100\text{ km/s}}{v}\right)^{1/4}\left(\frac{1\text{ kpc}}{r}\right)^{1/2} \tag{11.10}$$

This constraint is known as the Tremaire-Gunn limit. For example, in our galaxy $r \sim 10$ kpc and $v \sim 220$ km/s, giving the neutrino mass bound of $m_\nu \gtrsim 30$ eV, which is marginally compatible with the bound from neutrino-oscillation experiments and from the observed dark matter abundance of the whole universe (8.37) we found earlier. A more stringent bound is given by dwarf spheroidal galaxies whose typical masses are around $M \sim 10^6 M_\odot$. Those galaxies give the neutrino mass bounds of $m_\nu \gtrsim 300-500$eV that is highly incompatible with (8.37), thus suggesting that the Standard Model neutrinos cannot be a dark matter candidate.

11.3 Beyond the Standard Model dark matter candidates

11.3.1 Axions

A priori, there are good reasons to think that the QCD sector of the Standard Model violates CP. The most apparent consequence of the CP violation is expected to manifest in the form of a nonzero electric dipole moment of neutron, a quantity which has been measured to be zero to an extremely high accuracy. This puzzle of surprising absence of CP violation in QCD is known as the strong CP problem. A simple way to solve the strong CP problem is to introduce a new symmetry called the Peccei-Quinn (PQ) symmetry which, paired with appropriate cosmology, naturally drives the theory to a CP-conserving vacuum. Axion is the pseudo-Nambu-Goldstone boson generated by the spontaneous breaking of the PQ symmetry. In order for the axion to be a viable dark matter candidate, its mass must be in the range $m_a \sim 10^{-3} - 10^{-5}$ eV. Most axion detection experiments utilize the so-called Primakoff effect $a \to \gamma\gamma$ and its variants. In the Light-Shining-through-Walls (LSW) experiments, a light beam is directed towards a wall that is opaque to photons but transparent to axions. If light is detected on the other side of the wall then presumably some photons have converted into axions which easily went through the wall and converted back into photons. In the Axion Dark Matter eXperiment (ADMX), a microwave cavity is immersed in a strong magnetic field which, through a variant of the Primakoff effect, could stimulate axion conversions to RF-photons which can then be resonantly amplified. The Cern Axion Solar Telescope (CAST) is based on the same principle as that of ADMX but is focused on detecting axions emitted by the sun, instead of those making up the dark matter.

11.3.2 Sterile neutrinos

There are two big puzzles pertaining neutrinos. Every known type of fermion has been observed with both left- and right-handed chirality, except neutrinos; all the neutrinos that we have observed so far are left handed. Neutrinos are also observed to oscillate from one flavor to another, which indicates that the different neutrino flavours have different masses, which, in turn, means that at least two of the three flavours have non-zero, albeit very small, masses. The introduction of sterile neutrinos solves both puzzles and, possibly, also explains the dark matter. Sterile neutrinos are right-handed neutrinos which are expected to be present to complement the left-handed neutrinos that we have been observing. The addition of right-handed neutrinos allow us to add Dirac mass terms in the Lagrangian in order to explain the source of neutrino masses. The word "sterile" reflects the fact that the sterile neutrinos must be singlet representations of the Standard Model (or else we would have seen them by now), distinguishing themselves from the "active" Standard Model neutrinos. Having no charge of any kind, the sterile neutrinos could be their own antiparticles (majorana), and this allows us to add majorana mass terms for them. If we give large majorana masses to the sterile neutrinos, we could also explain the smallness of the masses of the Standard Model neutrinos naturally by the so-called seesaw mechanism. Very heavy sterile neutrinos are good dark matter candidates. The sterile neutrinos ν_R could decay through the process $\nu_R \to \nu_L \gamma$, which could be detected with x-ray satellites for the mass range that is compatible with them being the dark matter.

11.3.3 WIMPs

WIMPs stands for weakly interacting massive particles. It is a particularly compelling type of dark matter that deserves a special attention. Suppose that there is a heavy particle X which cannot decay to the Standard Model sector but can annihilate among themselves and assume that X and its antiparticle $\bar{X}$ are equally abundant. The abundance of X today is determined by its freezeout temperature, which happens when

$$\Gamma_X = \langle \sigma n v \rangle \sim H \tag{11.11}$$

If the mass and annihilation cross-section of X are M and $\sigma_{\text{ann}} = \sigma_0 / v$ respectively, its equilibrium abundance is given by

$$n_X = g_X \left(\frac{MT}{2\pi} \right)^{3/2} e^{-M/T} \tag{11.12}$$

Freeze out occurs when

$$\begin{aligned} \Gamma &\sim H \\ \sigma_0 n_X &\sim H \\ \sigma_0 g_X \left(\frac{MT}{2\pi} \right)^{3/2} e^{-M/T} &\sim \frac{T^2}{M_0} \end{aligned} \tag{11.13}$$

After taking the logarithm of the above, the approximate freeze-out temperature can be found by iteration

$$T_f \sim \frac{M}{\log\left(g_X M M_0 \sigma_0/(2\pi)^{3/2}\right)} \lesssim M \tag{11.14}$$

The present-day relic density can be calculated by rescaling the density at freeze out

$$n_X(t_0) = n_X(t_f)\frac{a^3(t_f)}{a_0^3} = n_X(t_f)\frac{s_0}{s(t_f)} = n_X(t_f)\frac{2T_\gamma^3 + 6\frac{7}{8}T_\nu^3}{g_*(t_f)T_f^3} \tag{11.15}$$

where the entropy today and at the moment of freeze out can be calculated as follows

$$s_0 = \frac{2\pi^2}{45}\left(2T_\gamma^3 + 6\frac{7}{8}T_\nu^3\right) \approx 2.8\times 10^{-3}\ \mathrm{cm}^{-3} \tag{11.16}$$

$$s(t_f) = \frac{2\pi^2}{45}g_*(T_f)T_f^3 \tag{11.17}$$

From last four equations and (4.11), we obtain the present relic abundance

$$\Omega_X = \frac{n_X(t_0)M}{\rho_c} = 3\times 10^{-10}\left(\frac{\mathrm{GeV}^2}{\sigma_0}\right)\frac{1}{\sqrt{g_*(T_f)}}\log\left(\frac{g_* M_0 M \sigma_0}{(2\pi)^{3/2}}\right) \tag{11.18}$$

As we can see, Ω_X depends most sensitively on σ_0, the only variable that is not suppressed by roots or logarithms. To get $\Omega_x \approx 0.3$, we need $\sigma_0 \sim 10^{-9}$ GeV^{-2}, which is incidentally close to the typical cross section of weak interactions

$$\sigma_W \sim G_F^2 E^2 \approx 10^{-10}\left(\frac{E}{\mathrm{GeV}}\right)^2 \mathrm{GeV}^{-2} \tag{11.19}$$

This interesting coincidence, known as the WIMP miracle, raised people's confidence on WIMP dark matter. Unfortunately, dark matter with cross-section of that order has been largely ruled out.

11.3.4 Massive compact objects (MACHOs)

MACHOs are compact non-luminous classical objects formed by baryonic or non-baryonic particles. Possible candidates include Q-balls (solitons whose stability are ensured by charge-conservation), wimpzillas (very heavy WIMPs which could be produced during the preheating), primordial black holes (black holes formed by the gravitational collapse of high density regions of the density fluctuations seeded by inflation). The detection method for MACHOs vary depending on the model. For example, Q-balls passing the Earth could be detected by the simultaneous observations of transient signals at all stations of the Global Network of Optical Magnetometers for Exotic Physics (GNOME).

11.3.5 Modified gravity

Another possibility for accounting the dark matter is by modifying the general theory of relativity at large scales. The only evidence for dark matter so far is through its gravitational interaction, which has only been tested at scales much smaller than the present Hubble radius, and so might not be correct at large distances. One attempt at modifying general relativity is the modified Newtonian dynamics (MOND). Normally, the non-relativistic equation of motion is given by the Newton's law $\mathbf{F} = m\mathbf{a}$. In MOND, the Newton's law is modified to $\mathbf{F} = m\frac{|a|}{a_0}\mathbf{a}$, where a_0 is the universal acceleration with the order of $a_0 \sim H_0$. The simplest version of MOND has been ruled out by bullet cluster observations. The possible generalizations of GR are not well constrained and most likely requires introducing new particles to fit the data.

11.4 Known properties of dark matter

The dark matter must be highly stable with a lifetime that is much longer than the age of the universe, or otherwise they would have decayed. If the dark matter were relatively light, $M \lesssim 1$ TeV, then it must be neutral and very weakly interacting, otherwise we would have easily detected its cosmic flux. In order to not break apart the large scale structures, the dark matter must not be relativistic at the onset of structure formation. If the dark matter are fermions, their mass must be below 400 eV (the Tremaine-Gunn bound we obtained earlier).

11.5 Unknown properties of dark matter

The mass of the particles making up the dark matter is so far unknown. The predictions of dark matter masses from different models vary wildly, they go from as light as 10^{-33} eV (stringy axions) to as heavy as 10^{24} GeV (supersymmetric Q-balls). Neither fermion nor boson has been ruled at as a possible dark matter candidate, i.e. the spin of the dark matter is not yet known. Since the presence of dark matter has been based entirely on its gravitational effects, not much is known about its non-gravitational interactions other than they must be very weak. For the same reason, the production mechanism of the dark matter and how they are embedded in the big picture of particle physics are still subjects of speculation.

> Summary:
> Various independent observations suggest the presence of non-luminous non-absorbing substance, called dark matter, which makes up around 27% of the mass of the universe. The Standard Model does not seem to provide any acceptable candidate for dark matter. Possible beyond the standard model candidates for dark matter include: axions, sterile neutrinos, WIMPs, MACHOs, and modified gravity.

VI

Inflation

Lecture 12

horizon problem; flatness problem

12.1 Problems of the ΛCDM model and inflation as their solution

The theory of inflation is best understood as a paradigm rather than a specific model built from well-established physics. The paradigm can be defined as the acceptance of the conjecture that the universe, at its early stage, underwent a period of accelerated expansion to which various features of the observable universe can be attributed. Including such a period solves the horizon, flatness, and unwanted relics problem classically and at the same time provides the seeds of the large scale structures quantum mechanically. To date, no known alternative model can explain the same features of the universe as well as inflation does. Although the exact mechanism of inflation remains speculative, the strength of inflation is that it is a very natural phenomenon whose explanatory success does not depend sensitively on the details of its model. In addition, the possibility that there are many problems with a common explanation is highly encouraging.

12.1.1 Horizon problem

In a Minkowski spacetime, two events separated by $\Delta\ell$ spatially and Δt temporally are causally independent if $\Delta\ell > c\Delta t$; they cannot affect one another even in principle. How do we get an equivalent statement in an expanding universe? In short, the expansion of the universe allows light to travel more distance than it could on its own. The distance ℓ of a photon relative to a specified comoving point it is moving away radially from changes as

$$\frac{d\ell}{dt} = c + \frac{\dot{a}}{a}\ell \tag{12.1}$$

which can be solved easily by rewriting it as

$$\frac{d}{dt}\left(\frac{\ell}{a}\right) = \frac{c}{a} \tag{12.2}$$

Integrating the above equation from an initial time t_0 to an arbitrary time t gives us the total distance $\ell_H(t)$ a photon can travel between the time interval

$$\ell_H(t) = c\int_{t_0}^{t} \frac{a(t)}{a(t')}dt' \tag{12.3}$$

Recall that in a radiation dominated (RD) universe $a(t) \sim t^{1/2}$ and in a matter dominated (MD) universe $a(t) \sim t^{2/3}$. In those cases, $\ell_H(t)$ converges when we set $t_0 = 0$ (the initial singularity)

$$\ell_H(t) = \begin{cases} 2t, & \text{RD} \\ 3t, & \text{MD} \end{cases} \tag{12.4}$$

The finiteness of ℓ_H tells us that only points separated by distances less than ℓ_H have a chance to be in causal contact. In other words, ℓ_H sets the radius of the *particle horizon* beyond which a particle has no influence over and cannot be influenced by. The fundamental reason behind the existence of particle horizon, i.e. the finiteness of ℓ_H, is our assumption that the universe has a beginning (the big bang). If the universe has no beginning, then the past light cones of various parts of the universe can extend infinitely far into the past and must overlap at some point.

The temperature profile of the CMB shows that the temperature of the universe at the point of photon decoupling was uniform within 1 part in 10^5, suggesting that the whole patch of the observable universe were once in thermal equilibrium. Peculiarly, the backward extrapolation of the ΛCDM model for as long as the age of the universe indicates that most CMB spots apparently have non-overlapping past light cones and thus were never in causal contact. To see this explicitly, the size of the particle horizon $\ell_H(t_d)$ when photons decoupled is given by (12.4)

$$\ell_H(t_d) = 3t_d \tag{12.5}$$

with $t_d = 5 \times 10^5$ years. Due to the expansion of the universe, its size today is bigger by a factor of a_0/a_d

$$\ell_H(t_d)\frac{a_0}{a_d} = 3t_d\left(\frac{t_0}{t_d}\right)^{2/3} \tag{12.6}$$

which spans an angular size of

$$\theta_H = \frac{\ell_H(t_d)a_0/a_d}{\ell(t_0)} = \frac{3t_d}{3t_0}\left(\frac{t_0}{t_d}\right)^{2/3} = \left(\frac{t_d}{t_0}\right)^{1/3} \tag{12.7}$$

on the sky. Plugging $t_0 \approx 15 \times 10^9$ years, we get $\theta_H \sim 1/30 \sim 2^\circ$. The number of causally separated regions within our horizon today is roughly given by

$$\left[\frac{\ell(t_0)}{\ell_H(t_d)a_0/a_d}\right]^3 \sim 3 \times 10^4 \tag{12.8}$$

We expect temperatures in these regions to be uncorrelated, but the CMB was found to be isotropic to a high accuracy (1 part in 10^5) over angular scales much larger than one degree. This so-called *horizon problem* calls for an explanation for the physical mechanism that was responsible for coordinating the temperatures in the seemingly causally-disconnected parts of the universe.

As we will soon see, the horizon problem can be solved if we conjecture a period in the early universe where the universe was expanding at an accelerated rate, i.e. a period of inflation. The existence of such a period implies that the early universe was much smaller than its expected size according to the backward extrapolation of ΛCDM model. Hence, the different parts of the present observable universe could once be in causal contact, and the observed homogeneity and isotropy is nothing remarkable, but rather an expected outcome.

12.2 Flatness problem

The Friedmann equation (4.9) can be rearranged as

$$\frac{\rho_{\text{tot}}}{\rho_c} - 1 = \frac{K}{a^2H^2} \tag{12.9}$$

where $\rho_{\text{tot}} = \rho_\Lambda + \rho_m + \rho_\gamma$. The quantity K/a^2H^2 can be taken as the curvature abundance ρ_K/ρ_c at an arbitrary time, which can be taken as a measure of the curvature of the universe. In the RD or MD epoch, we have $H \sim 1/t$ and $a \sim t^\alpha$ with $\alpha = 1/2$ (RD) and $\alpha = 2/3$ (MD). Hence, the above measure of curvature evolves as

$$\frac{K}{a^2H^2} \sim t^{2(1-\alpha)} \tag{12.10}$$

i.e. increasing with time. At the present epoch, observations tell us that the universe is flat to a high accuracy

$$\left.\frac{K}{a^2H^2}\right|_{t=t_0} \lesssim 10^{-2} \tag{12.11}$$

The fact that the current universe appears to be flat is surprising because in the ΛCDM model any deviation from the flatness would increase with time. Apparently, in order to have $K/a^2H^2 \lesssim 10^{-2}$ now, K/a^2H^2 must be fine-tuned to extremely small numbers in the past. For instance, we require $K/a^2H^2 \sim 10^{-15}$ at the time of nucleosynthesis in order to be consistent with the present observations. This puzzle of surprising flatness of the universe is dubbed the *flatness problem*.

A plausible solution to the problem shows itself when we realize that if for some reason the α in (12.10) is effectively less than 1, a deviation from flatness would decrease rather than increase with time. Again, this can be achieved by introducing a period of inflation. If the universe underwent a sufficiently long period of inflation before the radiation domination, K/a^2H^2 could be driven to an exponentially small value that the amplification it receives since the radiation domination era is not enough to make it significant today. To get a better grasp of the idea, let us see how it works explicitly.

Suppose that the universe was dominated by the vacuum energy at some point before the baryogenesis, BBN, and photon decoupling. As we found earlier, a vacuum-energy dominated universe expands as

$$a(t) = a_0 \exp\left[H(t-t_0)\right] \tag{12.12}$$

with $H = \sqrt{8\pi G\rho_\Lambda/3} = \text{const}$, where ρ_Λ is the vacuum energy density dominating the energy budget of the universe during the period of inflation. Then, at $t = t_1$ the vacuum-energy dominated epoch ends with a transition to radiation-domination epoch. After that, the scale factor continues to evolve as

$$a(t) = a_0 \exp\left[H(t_1-t_0)\right]\left(\frac{t}{t_1}\right)^{1/2} \tag{12.13}$$

Hence, using (12.3) we can calculate the horizon size at photon decoupling as

$$\begin{aligned}\ell_H(t_d) &= \int_{t_0}^{t_1} dt' \exp\left[H(t-t')\right] + \int_{t_1}^{t_d} dt' \left(\frac{t}{t'}\right)^{1/2} \\ &= \frac{1}{H}\exp\left[H(t_1-t_0)\right] + 2(t_d - t_1) \\ &\sim \frac{1}{H}\exp\left[H(t_1-t_0)\right]\end{aligned} \tag{12.14}$$

where in the last step we have assumed that the period of inflation was sufficiently long that $H(t_1-t_0) \gg 1$ and $H^{-1}\exp[H(t_1-t_0)] \gg 2(t_d-t_1)$. If, for example, $\rho_\Lambda \sim \left(10^{15}\text{ GeV}\right)^4$ (GUT scale), the horizon problem is solved if the particle horizon at the time of photon decoupling as seen today $\ell_H(t_d)a_0/a_d$ is at least as large as the present Hubble radius H_0^{-1}. The latter requirement gives a lower bound on the number of e-foldings during inflation $H(t_1-t_0) \gtrsim 65$. The flatness problem is also automatically solved since

$$\left.\frac{K}{a^2H^2}\right|_{t=t_0} \propto K\exp[-2H(t_1-t_0)] \tag{12.15}$$

becomes exponentially suppressed if $H(t_1-t_0) \gg 1$. The time-dependence of K/a^2H^2 is, of course, more complicated than displayed above, but the point is that as long as $H(t_1-t_0) \gg 1$ the exponential suppression will most likely be the dominant factor.

Summary:
Unsurprisingly, such a simple model as the ΛCDM model has a number of shortcomings. Among others, (1) the model does not provide a physical mechanism for coordinating the different parts of the universe that seem to be causally disconnected (if we extrapolate ΛCDM backwards) to within 1 part in 10^5 (horizon problem) and (2) it does not explain why the current universe is very flat, especially considering that within the ΛCDM model any deviation from flatness in the past would grow (flatness problem). Both problems can solved by introducing an era of accelerated expansion, dubbed inflation. The horizon problem is solved because a sufficiently long period of inflation implies that the universe at its earliest stage was much smaller than suggested by the ΛCDM model, allowing the past light cones of the different parts of the current universe to overlap, and the flatness problem is solved because inflation drove the universe exponentially close to flatness and for that reason the deviation from flatness stays small despite being amplified afterwards.

Lecture 13

chaotic inflation; density perturbation from inflation

13.1 Chaotic inflation

Inflation is a very natural phenomenon that could occur in many typical circumstances. So, even if it were not to explain anything, one should still seriously consider the possibility of its occurrence. To demonstrate how easy it is to give rise to a period of inflation, here we consider arguably the simplest model of inflation, chaotic inflation. Consider a free scalar field with the action

$$S=\int d^4x\sqrt{-g}\left[\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi-\frac{1}{2}m^2\phi^2\right] \tag{13.1}$$

with $m \ll M_p$ and "chaotic" initial condition at the Planck scale $t\sim M_P^{-1}\sim 10^{-43}$ s, by which we mean that the initial energy density is given by

$$\rho=\frac{1}{2}\dot{\phi}^2+\frac{1}{2}(\nabla\phi)^2+\frac{1}{2}m^2\phi^2\sim M_P^4 \tag{13.2}$$

Inflation is expected to happen in regions where the potential energy $V(\phi)=m^2\phi^2/2$ dominates the energy density

$$m^2\phi^2 \gg \dot{\phi}^2\sim(\nabla\phi)^2\sim M_P^4 \tag{13.3}$$

That would require

$$\phi\sim\frac{M_P^2}{m}\gg M_P \tag{13.4}$$

Let us now focus on those regions. For simplicity, in what follows we will assume that the scalar field is homogeneous. Note that this assumption is not a strong one. If the scalar field was not initially homogeneous, there are always regions in which the ϕ field satisfies (13.4). Even if these regions are initially extremely small and rare, they would inflate exponentially and soon dominate the space. The ϕ field obeys the classical equation of motion

$$\ddot{\phi}+3H\dot{\phi}+m^2\phi=0 \tag{13.5}$$

which is an equation for damped harmonic oscillator, where

$$H^2 = \frac{8\pi G}{3}\left[\frac{1}{2}\dot{\phi}^2 + \frac{1}{2}m^2\phi^2\right] \approx \frac{4\pi m^2\phi^2}{3M_P^2} \tag{13.6}$$

where we have assumed that the potential energy dominates the energy density. The motion of the ϕ field is overdamped (slow-roll) if

$$H \gg \frac{\ddot{\phi}}{\dot{\phi}} \tag{13.7}$$

In that case, the acceleration term in (13.5) can be neglected, leaving us with

$$\frac{\sqrt{12\pi}m\phi\dot{\phi}}{M_P} + m^2\phi = 0 \tag{13.8}$$

The solution is a constant velocity motion

$$\phi = \phi_0 - \frac{mM_P}{\sqrt{12\pi}}t \tag{13.9}$$

where ϕ_0 must satisfy $\phi_0 \gtrsim M_P^2/m$. The inflation ends when the the potential energy $m^2\phi^2$ becomes comparable to the kinetic energy $\dot{\phi}^2 \sim m^2 M_P^2$, that is when

$$\phi \sim M_P \tag{13.10}$$

or

$$t_* \sim \frac{M_P}{m^2} \tag{13.11}$$

By the time the inflation ends, the universe has inflated by a factor of

$$\exp(Ht_*) \sim \exp\left(\frac{M_P^2}{m^2}\right) \tag{13.12}$$

If one takes inflation to be the correct theory, there is still a “dark age" to be filled between the end of inflation and the beginning of big bang nucleosynthesis. At the end of chaotic inflation, the energy that drove inflation is in the form of the potential energy of the ϕ field. In order to switch to the hot Big Bang cosmology, at some point, this energy density is expected to decay to produce radiation particles, heating up the cold, diluted universe. This transition process is known as *reheating*. Note that the term reheating is a potential misnomer implying that the universe was once hot before the inflation. As a matter of fact, we do not have enough theoretical and observational knowledge to make reliable inference about the epoch before inflation.

Physicists’ creativity at cooking up theories that give rise to inflation has given birth to assorted models of inflation, e.g. hybrid inflation, power-law inflation, new inflation, old inflation, etc. These models are beyond the scope of these lectures. Since the constraints on the mechanism of inflation are still pretty loose, there is no point in stressing too much on any of these models.

13.2 Density perturbation from inflation

The theory of inflation was originally motivated by its ability to solve the horizon problem and flatness problem simultaneously via its classical dynamics. Later, physicists realized that, as a bonus, the quantum aspect of the same idea could provide the appropriate initial conditions for the primordial density perturbation. We have a pretty good understanding of how the initial density perturbation evolves to the profile captured by CMB, provided that the initial perturbation is scale-invariant, Gaussian, and adiabatic; the most general form of density perturbation can be obtained from a linear combination of isocurvature perturbation (fluctuation in relative densities of different species) and adiabatic density perturbation (fluctuation in the total density without interchanges between the densities of different species). Remarkably, generic models of inflation could prepare primordial density perturbation with all these features granted that the accelerated expansion occurs at nearly constant H.

Qualitatively, inflation seeds the density perturbations in the universe in the following way. Quantum fluctuations $\delta\phi$ in the inflaton field results in spatial variation in the duration of inflation δt_{inf}, which, in turn, gives rise to a curvature perturbation ζ. By an appropriate choice of coordinates, the curvature perturbation ζ can be re-interpreted as an adiabatic density perturbation $\delta\rho$ on top of the homogeneous background density ρ.

Now, consider the Fourier transform δ_k of the density perturbation $\delta\rho/\rho$

$$\frac{\delta\rho}{\rho} = \frac{1}{(2\pi)^3}\int d^3\mathbf{k}\delta_k \exp\left(-i\mathbf{k}.\mathbf{x}\right) \tag{13.13}$$

where k is the comoving wavenumber. By definition, $\langle\delta_k\rangle = \langle\delta\rho\rangle = 0$, but the following quantity (correlation function) is in general non-zero

$$\left\langle\frac{\delta\rho(\mathbf{x})}{\rho},\frac{\delta\rho(\mathbf{y})}{\rho}\right\rangle = \frac{1}{(2\pi)^6}\int d^3p d^3q \exp\left(-i\mathbf{p}.\mathbf{x} - i\mathbf{q}.\mathbf{y}\right)\langle\delta_{\mathbf{p}}\delta_{\mathbf{q}}\rangle \tag{13.14}$$

Translation invariance or momentum conservation implies that

$$\langle\delta_{\mathbf{p}}\delta_{\mathbf{q}}\rangle = (2\pi)^3\delta\left(\mathbf{p}+\mathbf{q}\right)|\sigma_{\mathbf{p}}|^2 \tag{13.15}$$

and consequently the correlation function reduces to

$$\left\langle\frac{\delta\rho(\mathbf{x})}{\rho},\frac{\delta\rho(\mathbf{y})}{\rho}\right\rangle = \frac{1}{(2\pi)^3}\int d^3p \exp\left[-i\mathbf{p}.(\mathbf{x}-\mathbf{y})\right]|\sigma_{\mathbf{p}}|^2 \tag{13.16}$$

The newly introduced quantity $\sigma_{\mathbf{p}}$ is called the power spectrum. Inflation predicts that for $\mathbf{x}$ and $\mathbf{y}$ located within a causal domain $|\mathbf{x}-\mathbf{y}| \lesssim \ell_H(t_0)$, the correlation function is constant

$$\left\langle\frac{\delta\rho(\mathbf{x})}{\rho},\frac{\delta\rho(\mathbf{y})}{\rho}\right\rangle = \text{const} \tag{13.17}$$

Comparing this with (13.16) and (13.15), we find that the power spectrum must be of the form

$$|\sigma_{\mathbf{p}}|^2 \propto \frac{1}{p^3} \tag{13.18}$$

in order to give a constant correlation function. Such a scale-invariant power spectrum is called the Harrison-Zeldovich power spectrum. In reality, due the impossibility of having perfectly constant Hubble parameter during inflation, there is a slight tilt in the power spectrum

$$|\sigma_{\mathbf{p}}|^2 \propto \frac{1}{p^3}p^{n_S-1} \tag{13.19}$$

where $n_S \approx 1$ is called the scalar spectral index. In addition to scalar perturbations $\delta\rho/\rho$, inflation also predicts tensor perturbations. It is common to measure their size relative to the size of the scalar perturbation in terms of the tensor-to-scalar ratio

$$r = \frac{\rho_2}{\rho_0} \tag{13.20}$$

where ρ_2 is the energy of spin-2 fluctuations (associated to the gravitational waves created during inflation) and the energy in the scalar fluctuations.

To explain the observed CMB fluctuations, inflation gives as an input the form of the power spectrum $|\sigma_p|^2$, which then get modified by various "ordinary physics" effects, e.g. photon scattering, gravitational lensing, plasma waves, etc that we know how to incorporate. Once these effects are accounted, we obtain a prediction for the present form of temperature fluctuation $\left\langle \frac{\delta T(\mathbf{x})}{T}, \frac{\delta T(\mathbf{y})}{T} \right\rangle$ which we can compare with CMB measurements. The predicted form depends on many parameters, e.g. n_S, r, Ω_B, Ω_V, etc, enabling themselves to be determined by measuring the CMB temperature anisotropy. Since different physics took place at different length scales, the best way to disentangle the information carried by the CMB is to break it down into different length scales, which are projected onto the sky as different angular scales. Formally, we do that by expanding the temperature fluctuation $\Delta T/T$ in terms of the spherical harmonics (reminder: spherical harmonics are orthogonal sets of natural functions on which we can decompose any function on a sphere)

$$\frac{\delta T(\theta,\phi)}{T} = \sum_{\ell m} Y_{\ell m}(\theta,\phi) a_{\ell m} \tag{13.21}$$

where the multipole moment ℓ tells us how fast the function oscillates around the sphere and m tells us how the oscillating function is oriented around the sphere. So, each term in the ℓ expansion captures increasingly finer angular features of the temperature profile. When the same expansion is done on the correlation function, we get

$$\left\langle \frac{\delta T(\theta,\phi)}{T}, \frac{\delta T(\theta',\phi')}{T} \right\rangle = \sum_{\ell\ell'}\sum_{mm'} Y_{\ell m}(\theta,\phi) Y_{\ell' m'}(\theta',\phi') \langle a_{\ell m} a^*_{\ell' m'} \rangle \tag{13.22}$$

If the fluctuations are spherically symmetric, different ℓ's and different m's are uncorrelated, i.e.

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta_{\ell\ell'} \delta_{mm'} \tag{13.23}$$

The above relation can be understood as follows. For a given ℓ, all the $2\ell+1$ different $a_{\ell m}$'s are independently drawn from the same distribution whose variance is C_ℓ. Note that it is this C_ℓ that experimentalists like to plot when presenting the temperature-temperature power spectrum data. Since the universe allows us only a finite number of chance ($2\ell+1$ for each ℓ) to draw samples from the distribution, there is a limit to how accurately we can define C_ℓ's. This unavoidable source of uncertainty is famously known as the cosmic variance. The uncertainty in $a_{\ell m}$is higher for lower ℓ's because for smaller ℓ's we have smaller number of $a_{\ell m}$'s to average over. Using (13.23), we can write (13.22) as

$$\begin{aligned} \left\langle \frac{\delta T(\theta,\phi)}{T}, \frac{\delta T(\theta',\phi')}{T} \right\rangle &= \sum_{\ell\ell'}\sum_{mm'} Y_{\ell m}(\theta,\phi) Y_{\ell' m'}(\theta',\phi') C_\ell \delta_{\ell\ell'} \delta_{mm'} \\ &= \sum_\ell \frac{2\ell+1}{4\pi} P_\ell(\cos\alpha) C_\ell \end{aligned} \tag{13.24}$$

where P_ℓ's are the Legendre polynomials and α is the angle between the directions (θ,ϕ) and (θ',ϕ'). Noting that $P_\ell \propto (\cos\alpha)^\ell \sim \cos(\ell\alpha) + \ldots$, we see that P_ℓ captures angular fluctuations of size $\theta \sim \pi/2\ell$, or, in degrees $\theta \sim 100°/\ell$. It can be shown that C_ℓ's are related to the power spectrum Δ_T^2 via

$$\Delta_T^2 = \frac{\ell(\ell+1)}{2\pi} C_\ell T^2 \tag{13.25}$$

Summary:
Although the exact mechanism of inflation remains speculative, the strength of the theory of inflation is that it is a very natural phenomenon whose explanatory success does not depend sensitively on the details of its model. The chaotic inflation model demonstrates that a theory consisting of a free scalar field with the most generic initial conditions, i.e. with kinetic, gradient, and potential energies of order M_P^4, can give rise to inflation if the initial vacuum expectation value of the scalar field is much larger than M_P. Apart from solving the horizon problem and flatness problem, inflation also lays down the appropriate initial conditions (scale-invariant) for the primodial density perturbation which would grow to produce the observed large-scale structures.

VII Perturbations

Lecture 14

perturbations in an expanding universe; growth of perturbations in a static universe; Jeans mass

14.1 Perturbations in an expanding universe

We have seen in the previous lectures that at the point of recombination, the temperature fluctuations $\delta T/T$ in the universe are at the level of 10^{-5}. On the other hand, today's observations indicate that the universe is full of large scale structures, i.e. density perturbations with $\delta\rho/\rho \sim 1$. The aim in this lecture is to understand how to go from the former to the latter.

For the purpose of studying the perturbations in the universe, we can characterize the universe by its: gravitational potential $\phi(\mathbf{x})$, energy density $\rho(\mathbf{x})$, pressure $p(\mathbf{x})$, and velocity field $v(\mathbf{x})$. The evolutions of these quantities are dictated by: the Newton's law of gravity

$$\nabla^2\phi = 4\pi G\rho \tag{14.1}$$

the continuity equation

$$\frac{\partial\rho}{\partial t} + \nabla(\rho\mathbf{v}) = 0 \tag{14.2}$$

and the Euler equation

$$\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}.\nabla)\mathbf{v} = -\frac{1}{\rho}\nabla p - \nabla\phi \tag{14.3}$$

The least trivial of them is the Euler equation. It may seem complicated, but it is nothing but the Newton's 2nd law. To see this, consider a small volume element. The force acting on it is given by the sum of the force due to pressure gradient and the force due to gravitational potential gradient

$$\begin{aligned} F &= F_{\text{pressure}} + F_{\text{grav}} \\ &= -\int dV\,(\nabla p + \rho\nabla\phi) \end{aligned} \tag{14.4}$$

By the Newton's 2nd law, this force must be equal to $\int dV \rho d\mathbf{v}/dt$, i.e. the "$ma$" of $F = ma$. The velocity field $\mathbf{v}(\mathbf{x},\mathbf{y},\mathbf{z},t)$ at a specified point can change both due to its explicit time dependence and the effect of different parts of the fluid with different velocities continuously replacing one another

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial \mathbf{v}}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial \mathbf{v}}{\partial z}\frac{\partial z}{\partial t}$$
$$= \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}.\nabla)\mathbf{v} \tag{14.5}$$

Substituting the above expression into the $d\mathbf{v}/dt$ in $\int dV\rho d\mathbf{v}/dt$ and equating the result with (14.4) gives us the Euler equation we sought after.

The aforementioned treatment must at least reproduce the Friedmann equations in the case where the universe is homogeneous if we are to trust its results. Let us check if this is indeed the case. If we take the gravitational potential, energy density, and pressure to be homogeneous with values ϕ_0, ρ_0, and p_0, and take the fluid to be comoving with velocity field v_0 given by $\mathbf{v}_0 = H\mathbf{x}$, (14.1) reduces to

$$\nabla^2\phi = 4\pi G\rho_0 \tag{14.6}$$

which can be solved by the gravitational equivalent of the Gauss law

$$\phi_0 = \frac{2\pi G}{3}\rho_0 r^2 \tag{14.7}$$

The continuity equation (14.2) reduces to

$$\frac{\partial \rho_0}{\partial t} + \nabla(\rho_0 H\mathbf{x}) = 0$$
$$\frac{\partial \rho_0}{\partial t} + 3H\rho_0 = 0 \tag{14.8}$$

which is the familiar energy conservation equation (2.17) for a pressureless fluid, and finally (14.3) reduces to

$$\dot{H}x_i + (Hx_j\partial_j)Hx_i + \frac{4\pi G}{3}\rho_0 x_i = 0$$
$$\dot{H} + H^2 + \frac{4\pi G}{3}\rho_0 = 0 \tag{14.9}$$

which is one of the Friedmann equations, i.e. (2.12).

Now, we start to actually perturb the four fields

$$\rho(\mathbf{x},t) = \rho_0(t) + \delta\rho(\mathbf{x},t)$$
$$p(\mathbf{x},t) = p_0(t) + \delta p(\mathbf{x},t)$$
$$\mathbf{v}(\mathbf{x},t) = \mathbf{v}_0(t) + \delta\mathbf{v}(\mathbf{x},t)$$
$$\phi(\mathbf{x},t) = \phi_0(t) + \delta\phi(\mathbf{x},t) \tag{14.10}$$

Inserting them to (14.1), (14.2), and (14.3) and keeping only the terms linear in the perturbations, we get

$$\nabla^2\delta\phi = 4\pi G\delta\rho \tag{14.11}$$
$$\frac{\partial \delta\rho}{\partial t} + \nabla(\delta\rho\mathbf{v}_0) + \nabla(\rho_0\delta\mathbf{v}) = 0 \tag{14.12}$$
$$\frac{\delta \mathbf{v}}{\partial t} + (\mathbf{v}_0.\nabla)\delta\mathbf{v} + (\delta\mathbf{v}.\nabla)\mathbf{v}_0 + \frac{1}{\rho_0}\nabla\delta p + \nabla\delta\phi = 0 \tag{14.13}$$

If the tiny fluctuations at the point of recombination are to yield the large scale structures that we see today, the above set of equations should tell us that some of them must grow with time. We will show this in the next section.

14.2 Growth of perturbations in a static universe

Perturbations in the universe can grow in a static universe. The expansion of the universe only complicates the matter. So, to start with, we will assume that the universe is static, i.e. $\mathbf{v}_0 = H\mathbf{x} = 0$, and so (14.11), (14.12), and (14.13) reduce to

$$\nabla^2 \delta\phi = 4\pi G \delta\rho \tag{14.14}$$

$$\frac{\partial \delta\rho}{\partial t} + \nabla(\rho_0 \delta\mathbf{v}) = 0 \tag{14.15}$$

$$\frac{\partial \delta\mathbf{v}}{\partial t} + \frac{1}{\rho_0}\nabla\delta p + \nabla\delta\phi = 0 \tag{14.16}$$

The Euler equation (14.16) can be rewritten in terms of the speed of sound $v_S^2 = \delta p/\delta\rho$ as

$$\frac{\partial \delta\mathbf{v}}{\partial t} + \frac{v_S^2}{\rho_0}\nabla\delta\rho + \nabla\delta\phi = 0 \tag{14.17}$$

Next, to relate the continuity equation (14.15) with the above equation, we take the time derivative of the former

$$\begin{aligned}
\frac{\partial^2 \delta\rho}{\partial t^2} + \rho_0 \nabla\left(\frac{\partial \delta\mathbf{v}}{\partial t}\right) &= 0 \\
\frac{\partial^2 \delta\rho}{\partial t^2} + \rho_0 \nabla\left(-v_S^2 \frac{1}{\rho_0}\nabla\delta\rho - \nabla\delta\phi\right) &= 0 \\
\frac{\partial^2 \delta\rho}{\partial t^2} - v_S^2 \nabla^2 \delta\rho - \rho_0 4\pi G \delta\rho &= 0
\end{aligned} \tag{14.18}$$

From the first to second line (14.17) was used and from the second to third line (14.14) was used. Injecting the ansatz $\rho \propto \exp(i\mathbf{k}.\mathbf{x} - i\omega t)$ into the last equation, we get

$$\omega^2 = v_S^2 k^2 - 4\pi G \rho_0 \tag{14.19}$$

Thus, we learn that perturbations with $k^2 > \sqrt{4\pi G\rho_0}/v_S$ are stable and those with $k^2 < \sqrt{4\pi G\rho_0}/v_S$ grow exponentially with time. It is the latter modes which went on to become the structures that we observe. The wavelength separating stable and unstable wavelengths

$$\lambda_J = \frac{2\pi v_S}{\sqrt{4\pi G\rho_0}} \tag{14.20}$$

is known as the Jeans wavelength.

Now we turn to the question: what happens to the perturbations in an expanding universe? In a matter dominated universe, the stability/instability condition remains unchanged, but instead of growing exponentially the unstable modes grow according to a power law $\delta\rho \propto t^{2/3}$ due to the Hubble friction. In a radiation dominated universe, on the other hand, even the unstable modes do not grow significantly; they grow logarithmically $\delta\rho \propto \log t$. Therefore, we conclude that structure formations must have occurred during the matter domination epoch.

In ways analogous to the Jeans wavelength, we can define a number of other quantities. In particular, the Jeans mass M_J is defined as the total mass inside a sphere of radius equals to the Jeans wavelength λ_J (14.20)

$$M_J = \frac{4\pi}{3}\lambda_J^3 \rho_0 = \frac{4\pi}{3}\rho_0 \left(\frac{\pi v_S^2}{G\rho_0}\right)^{3/2} \tag{14.21}$$

The Jeans mass M_J tells us the minimum amount of mass within a Jeans volume (volume of a sphere with radius equals to the jeans wavelength) above which the density grows. In the radiation domination epoch, $p = \rho/3$, and so $v_S^2 = \partial p/\partial\rho = 1/3$, which means

$$M_J = \frac{4\pi}{3}\frac{M_P^3}{\sqrt{\rho_0}} \sim \frac{M_P^3}{T^2} \sim 10^{17} M_\odot \tag{14.22}$$

where we have set T to its value at the recombination. This M_J is huge in comparison to the mass of a typical galaxy, $10^4 M_\odot$, meaning that a typical galaxy would not collapse gravitationally.

14.3 Classification of dark matter: cold, warm, hot (not presented)

The dark matter can be classified based on its ability to flatten out perturbations. A good measure of this feature is the *free-steaming scale* λ_{FS}, which is defined as the distance a dark matter particle would have traveled in t_E (the age of the universe at the matter-radiation equality, at which point perturbations start to grow)

$$\lambda_{\mathrm{FS}} = v t_E \tag{14.23}$$

Structures with wavelengths smaller than λ_{FS} cannot grow due to the free-streaming of dark matter. We can then define the

$$M_{\mathrm{FS}} = \frac{4\pi}{3}\lambda_{\mathrm{FS}}^3 \tag{14.24}$$

and classify dark matter in the following way

$$\begin{aligned} 10^{14} M_\odot < M_{\mathrm{FS}} & \qquad \text{hot dark matter} \\ 10^{5} M_\odot < M_{\mathrm{FS}} < 10^{14} M_\odot & \qquad \text{warm dark matter} \\ M_{\mathrm{FS}} < 10^{5} M_\odot & \qquad \text{cold dark matter} \end{aligned}$$

For sterile neutrino, we find

$$M_{\mathrm{FS}} = 2.6 \times 10^{11} M_\odot \Omega_m h^2 \left(\frac{1\ \mathrm{GeV}}{m_\nu}\right)^3 \left(\frac{g/T}{3.15}\right)^3 \tag{14.25}$$

14.4 Overall summary of cosmology (not presented)

Now that we have discussed at length about different parts of cosmology, let us briefly recount the history of the universe according to our current best understanding

Table 14.1: Cosmic history.

Temperature / redshift / time	key event
$T > 10^{14}$ GeV	inflation
$T \sim 10^2 - 10^{14}$ GeV	baryogenesis
$T \sim 1$ MeV / $t \sim 1$ s – a few mins	Big Bang nucleosynthesis
$T \approx 0.7$ eV / $z \approx 3000$ / $t \approx 8 \times 10^4$ years	radiation-matter equality
$T \approx 0.25$ eV, $z \approx 1100$ / $t \approx 2.7 \times 10^5$ years	recombination and photon decoupling
today / $t \approx 14 \times 10^9$ years	matter - dark energy equality

Summary:
To study the growth of perturbations in an expanding universe, it is sufficient to keep track of its the gravitational potential, energy density, pressure, and velocity field, which evolve according to the Newton's law of gravity, continuity equation, and Euler equation. If the universe is static, we found there is a critical wavenumber above which the perturbation is stable and below which the perturbation grows exponentially. As far as the ΛCDM model is concerned, perturbations can only grow in the matter domination epoch. Hence, it follows that structure formation must take place during the period of matter domination.

Bibliography

[1] Dmitry S. Gorbunov and Valery A. Rubakov. *Introduction to the theory of the early universe: Cosmological perturbations and inflationary theory*. 2011. DOI: 10.1142/7874. URL: http://www.DESY.eblib.com/patron/FullRecord.aspx?p=737613.

[2] Edward W. Kolb and Michael S. Turner. “The Early Universe”. In: *Front. Phys.* 69 (1990), pages 1–547.

[3] V. Mukhanov. *Physical Foundations of Cosmology*. Oxford: Cambridge University Press, 2005. ISBN: 0521563984, 9780521563987. URL: http://www-spires.fnal.gov/spires/find/books/www?cl=QB981.M89::2005.

[4] Steven Weinberg. *Gravitation and Cosmology*. New York: John Wiley and Sons, 1972. ISBN: 0471925675, 9780471925675. URL: http://www-spires.fnal.gov/spires/find/books/www?cl=QC6.W431.

[5] Steven Weinberg. *Cosmology*. 2008. ISBN: 9780198526827. URL: http://www.oup.com/uk/catalogue/?ci=9780198526827.