ASTR/PHY 395 - Cosmology

Lecture notes

Timm Wrase

Abstract

These lecture notes will provide an introduction to cosmology, i.e., the large-scale evolution of our universe from the big bang until today and into the far future. They are roughly separated into three parts

Part I: The evolution of our universe: This part provides the background in geometry that is needed to describe the evolution of our universe from early times into the far distant future. Topics covered are: the Hubble expansion, the Friedmann equations, the content of our universe, its age and potential future fate, particle and event horizons and the cosmic microwave background (CMB).

Part II: The thermal universe: This part discusses the evolution of particles in the early universe up to a time of 380,000 years after the big bang. Topics covered are: the standard model of particle physics, baryogenesis/leptogenesis, equilibrium thermodynamics, particle freeze-out, the cosmic neutrino background, dark matter, big bang nucleosynthesis and recombination.

Part III: Inflation in the very early universe: This part provides an introduction to the theory of inflation, which is a conjectured period of accelerated expansion during our very early universe. Topics covered are: the horizon, flatness and monopole problems, inflation from a scalar field, slow-roll inflation, small field models of inflation, large field models of inflation, experimental constraints on inflationary models, beyond slow-roll, reheating, the inhomogeneous universe and quantum fluctuations as seed for our galaxies.

Contents

Part I - The evolution of our universe

1 The expanding universe

In this section we will learn about the discovery of the expansion of our universe, as well as the fact that the universe is homogeneous and isotropic on scales larger than a few *Mpc* (megaparsec). This allows us to derive a set of simple equations, the so called Friedmann equations, from general relativity. These equations play a central role in describing the evolution of our universe.

1.1 The 'Hubble' expansion

When Einstein first wrote down his theory of general relativity in 1915, he was convinced (like most other people) that our universe is static, i.e., the universe as a whole doesn't change in time. However, in 1929 Hubble was able to determine the distances and relative velocities of other galaxies by observing Cepheid variables which led him to a very different universe. In order to understand this, it is useful to first review distance measurements in astrophysics.

1.1.1 Astronomical unit and parsec

There are a variety of different units used in cosmology and astrophysics. One standard unit in astrophysics is the average distance between the earth and the sun which by definition is one astronomical unit $1au \approx 150 \times 10^6 km = 1.5 \times 10^{11} m$. In cosmology we are interested in larger scales and will mostly use the parsec (pc). The definition of the parsec involves the apparent parallax motion of near stars that is due to the earth's motion around the sun, see figure 1.

From simple trigonometry we find

$$1pc = \frac{1au}{\tan(1'')} \approx \frac{1au}{1''} = \frac{1au}{\frac{1}{60}\frac{1}{60}\frac{\pi}{180}} \approx 2 \times 10^5 au\,, \tag{1.1}$$

where we used that $\tan(1'') \approx 1''$. We can check that one parsec is roughly the distance light travels in three years $1pc \approx 3.3\,ly \approx 3.1 \times 10^{16} m$.

By measuring the parallax angle, astronomer can determine the distance of objects that are not too far away. This leads to interesting discoveries that can then be used to determine the distances of much further objects. In particular by studying nearby so-called Cepheid variables, astronomers found that these stars pulsate radially with a well-defined relation between their pulsation period and luminosity L (the total 'light' emitted by the star per time). By knowing this relation and the pulsation period we can therefore obtain the stars luminosity L. The observed flux F then directly gives us the distance to the Cepheid star since the observed flux decrease with the square of the distance d to the star. In particular we find

$$L = 4\pi\, d^2 F \qquad \Leftrightarrow \qquad d = \left(\frac{L}{4\pi F}\right)^{\frac{1}{2}}\,. \tag{1.2}$$

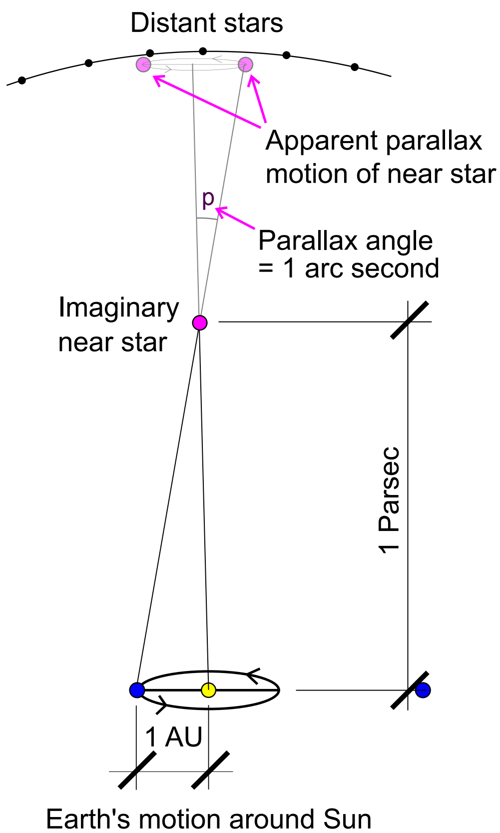


Figure 1: The distance to an imaginary star with a parallax angle of $1'' = 1$ arc second is one parsec (taken from Wikipedia).

Worked problem 1.1: Solar flux

The sun is 8 light minutes from earth and its luminosity is $L_\odot = 3.85 \times 10^{26}\ W$. Calculate the flux observed on earth.

Solution:

$$F = \frac{L_\odot}{4\pi d^2} = \frac{3.85 \times 10^{26}\ W}{4\pi(8 \cdot 1.8 \times 10^{10} m)^2} \approx 1480\ W/m^2\,. \tag{1.3}$$

The more exact value is 1373 W/m^2. On the earth surface this is reduced due to earth atmosphere by 15–80%.

Using these Cepheid variables one can determine the distances to many other objects, like for example type Ia supernovae. These are very bright explosions that turn out to have very consistent peak luminosities. So, knowing the peak luminosity we can fairly precisely determine the distance to any observed type Ia supernova in the universe. The total energy released in these explosions, that last a few days, is as large as the energy our sun releases in its lifetime of roughly $10^{10} yrs$. So, we can observe these supernovae very far out into the universe. The figure 2 shows how we can use parallax, Cepheid variables and type Ia supernovae to reach distances of roughly 10 *Mpc* to 10 *Gpc* that are relevant for cosmology.

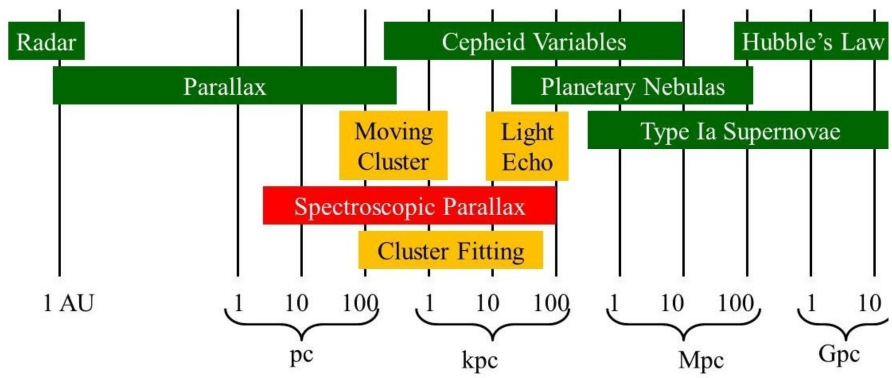


Figure 2: This figure shows the so-called cosmic distance ladder, i.e., different methods to measure distances in astronomy. (Credit: Tabitha Dillinger).

1.1.2 Hubble's discovery

Hubble studied these Cepheid variables in other galaxies and galaxy clusters and determined their distances using the above equation. In addition, he used the Doppler shift of the spectral lines in the star light to determine the relative velocities of these galaxies and galaxy clusters. This led him to the following plot

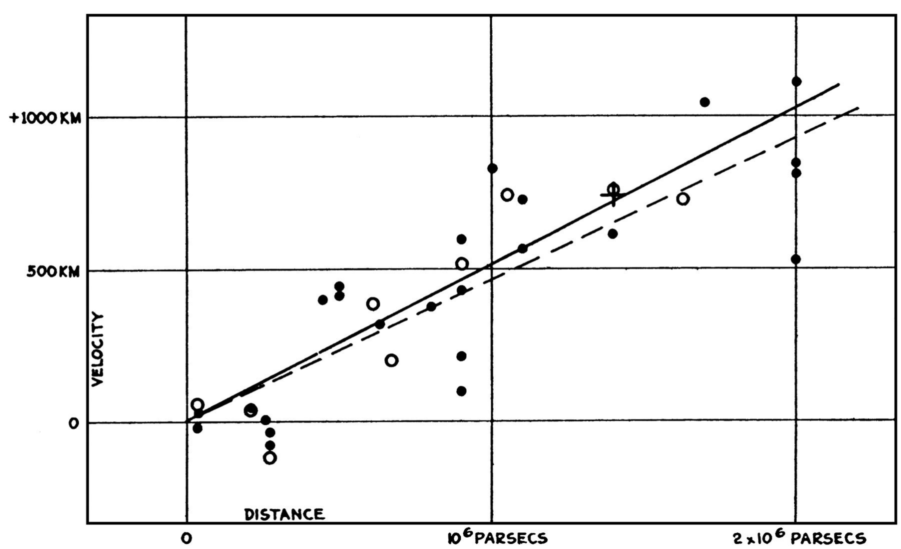


Figure 3: Velocity (in km/s) vs. distance (in parsecs) for galaxies (black dots) and galaxy clusters (circle). The solid line represents a best straight-line fit to the black dots and the dashed line to the circles.

It follows from Hubble's data that the further a galaxy is away from us, the faster it is moving away from us. This observation has been substantially improved over the year, as is shown for example in figure 4.

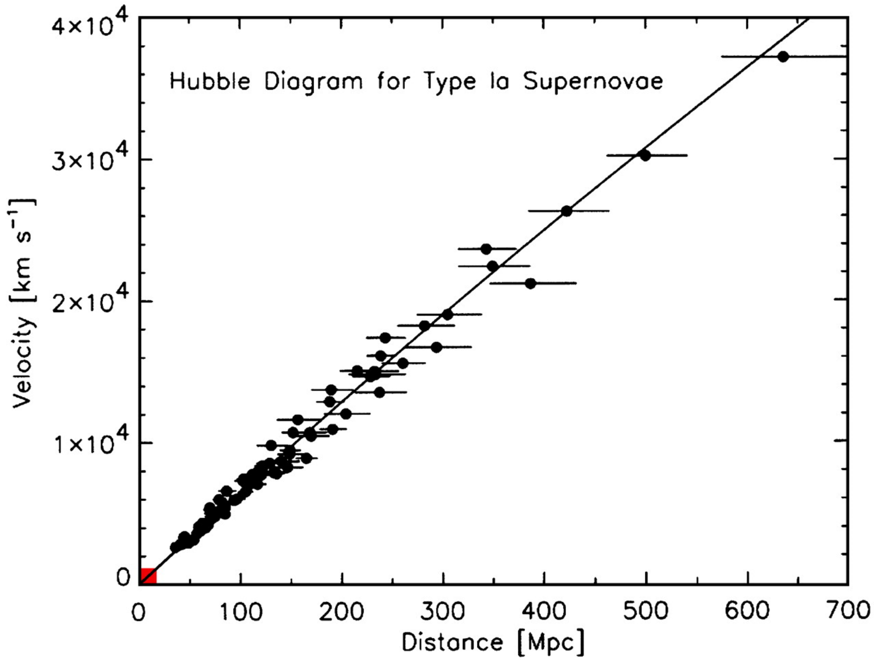


Figure 4: Velocity (in km/s) vs. distance (in Mpc) for Type Ia supernovae (another class of 'standard candles' that allows us to determine distances accurately).

Hubble's original observation is inconsistent with a static universe and instead requires us to consider a universe in which space itself is expanding. This is depicted in figure 5.

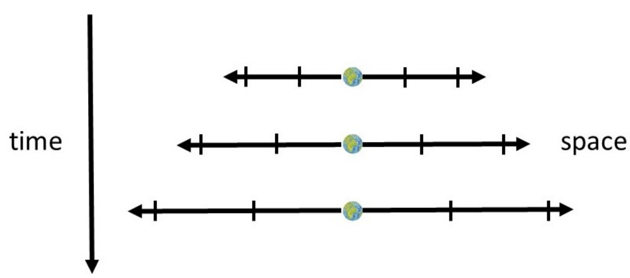


Figure 5: A one dimensional universe in which space itself is expanding. This leads to a linear relation between the relative distance and relative velocity of the earth and any other object in this universe.

Note however, that this does not make the earth or us special in anyway. Any other point in space will observe exactly the same 'Hubble' expansion of the universe, as is shown in figure 6.

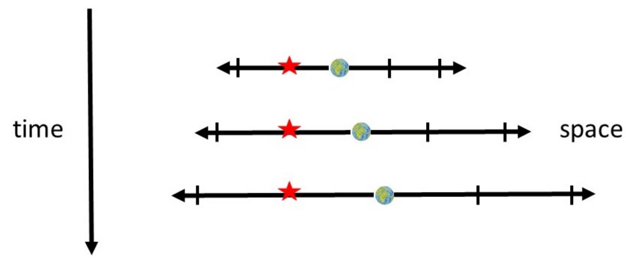


Figure 6: A one dimensional universe in which space itself is expanding. Any point in space will observe the same effect: distant objects are moving away with a velocity that is proportional to their distance.

Before we discuss the equations that describe such a universe, we discuss one more observational fact about our universe in the next subsection.

1.2 Isotropy and Homogeneity

Trying to describe the time evolution of the entire universe seems like a formidable task and one might wonder how this can be possibly done? In cosmology we are not interested in the details of the evolution on small scales like for example our solar system, but we would like to describe the origin, evolution and the ultimate fate of our universe. But even that seems intractable. Imagine a universe whose evolution is controlled by matter, i.e., at large scales by the evolution of the galaxies. This seems correspond to an N-body problem with N of the order of a few hundred billion ($N \sim 10^{11}$)!

Fortunately, our universe seems highly symmetric at scales larger than a few Mpc. Concretely, there is ample evidence that our universe looks the same in every direction, i.e., it is isotropic, and there are some indications that different locations in the universe allow for the same observation that we make, i.e., the universe is homogeneous. These two properties follow from the so called 'cosmological principle' that postulates that we do not occupy any special place in the (large scale) universe and other observers at any other place in the universe will observe the same properties of the universe. While this 'cosmological principle' originally was just a postulate, there is now ample observational data to support it.

Worked problem 1.2: Isotropy and Homogeneity

Draw a two-dimensional space that is homogeneous but not isotropic and one that is isotropic but not homogeneous.

Solution:

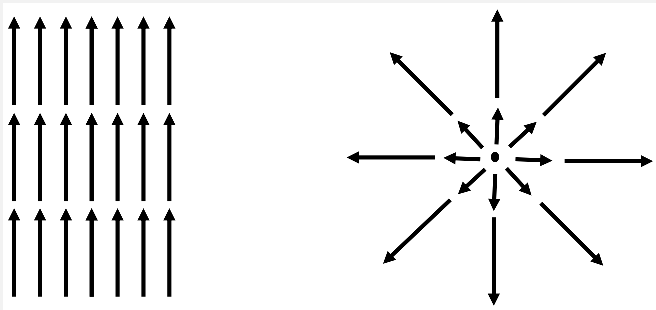


Figure 7: The picture on the left is homogeneous but not isotropic. The picture on the right is isotropic around the dot in the middle but not homogeneous.

The above picture would arise for example in settings with electric field lines between two charged plates (on the left) or an electric point charge (on the right).
Isotropy at each point does imply homogeneity, so the picture on the right cannot be modified in such a way that it looks isotropic at each point without also making it homogeneous.

A simple example of a 3-dimensional isotropic and homogeneous space is the flat space $\mathbb{R}^3$. The line element for this case is

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 \equiv dx_i^2 \,. \tag{1.4}$$

Another example of a space that is the same at every point and looks the same in every direction is the 3-sphere S^3 for which we can write the line element as[1]

$$ds^2 = dx_i^2 + dz^2 \,, \qquad x_1^2 + x_2^2 + x_3^2 + z^2 \equiv x_i^2 + z^2 = a^2 \,. \tag{1.5}$$

One can prove that the only other such space is given by a hyperspherical surface with negative curvature and line element

$$ds^2 = dx_i^2 - dz^2 \,, \qquad -x_i^2 + z^2 = a^2 \,. \tag{1.6}$$

By rescaling the x_i and z by a, we can write the last two as

$$ds^2 = a^2 \left(dx_i^2 \pm dz^2\right) \,, \qquad z^2 \pm x_i^2 = 1 \,. \tag{1.7}$$

[1]Note that the range of the x_i for a 3-sphere is not from negative to positive infinity but limited such that $x_i^2 \leq a^2$ or after the rescaling by a below one has $x_i^2 \leq 1$.

Differentiating $z^2 \pm x_i^2 = 1$ leads to $zdz = \mp(x_1dx_1 + x_2dx_2 + x_3dx_3) \equiv \mp x_idx_i$ and the line element

$$ds^2 = a^2\left(dx_i^2 \pm \frac{z^2dz^2}{z^2}\right) = a^2\left(dx_i^2 \pm \frac{(x_idx_i)^2}{1 \mp x_i^2}\right) . \tag{1.8}$$

Finally, we introduce the number $K \in \{-1, 0, 1\}$ and combine the three line elements (1.4), (1.5), (1.6) into one single equation

$$ds^2 = a^2\left(dx_i^2 + K\frac{(x_idx_i)^2}{1 - Kx_i^2}\right) . \tag{1.9}$$

Since we have chosen K to be dimensionless we have to choose the x_i to be dimensionless as well due to the denominator $1 - Kx_i^2$. Then the prefactor a needs to have the dimension of a length. In the above equation $K = 0$ corresponds to the flat space case and $K = \pm 1$ to the spherical and hyperspherical case. These three maximally symmetric three-dimensional spaces can be similarly defined in two space dimensions in which case we can picture them by embedding them into a three-dimensional space, as is shown in figure 8.

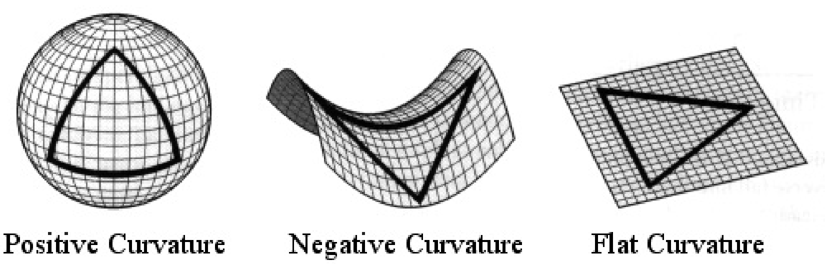


Figure 8: The three possible geometries of our universe.

As can be seen from the figure, the three spaces can be distinguished by measuring the three angels inside a triangle. For example, for a sphere one can start at the north pole with a 90° angle. These two sides each meet the equator at 90° angles and by choosing the third side to lie on the equator, we have constructed a triangle with total interior angles that add up to 270°! Generically one finds that for spherical geometries the three angles inside a triangle are larger than 180°, while for hyperspherical spaces they are smaller than 180°.

Worked problem 1.3: Calculating distances in curved space

Calculate the distance between the origin $x_i = 0$ and the point $x_1 = x_2 = 1/\sqrt{2}$, $x_3 = 0$ for $a = 1m$ and $K = 0, \pm 1$. Choose a path parametrized by $0 \leq \tau \leq 1/\sqrt{2}$ such that $x_1(\tau) = x_2(\tau) = \tau$.

Solution: The line element simplifies for $x_3 = 0$ and $x_1 = x_2 = \tau$ to

$$\begin{aligned} ds^2 &= a^2 \left(dx_1^2 + dx_2^2 + K \frac{(x_1 dx_1 + x_2 dx_2)^2}{1 - K(x_1^2 + x_2^2)} \right) \\ &= 2a^2 d\tau^2 \left(1 + K \frac{2\tau^2}{1 - 2K\tau^2} \right) . \end{aligned} \tag{1.10}$$

So, the distance is given by

$$\Delta s = \int_{s_1}^{s_2} ds = \sqrt{2}\, a \int_0^{1/\sqrt{2}} d\tau \sqrt{1 + K \frac{2\tau^2}{1 - 2K\tau^2}} . \tag{1.11}$$

For $K = 0$ we find

$$\Delta s = \sqrt{2}\, a \int_0^{1/\sqrt{2}} d\tau = a = 1\ m , \tag{1.12}$$

which is the expected result for flat space. For $K = +1$ we find

$$\Delta s = \sqrt{2}\, a \int_0^{1/\sqrt{2}} d\tau \frac{1}{\sqrt{1 - 2\tau^2}} = \frac{\pi}{2} a \approx 1.6\ m , \tag{1.13}$$

and for $K = -1$ we get

$$\Delta s = \sqrt{2}\, a \int_0^{1/\sqrt{2}} d\tau \frac{1}{\sqrt{1 + 2\tau^2}} = \text{arcsinh}\,(1)\ a \approx .88\ m . \tag{1.14}$$

Now that we have understood the spatial part of our universe, we can extend the line element to include also time and write (note that we will set the speed of light $c \approx 3 \times 10^8 m/s$ equal to 1 so that $1s \approx 3 \times 10^8 m$, see the next subsection for details)

$$ds^2 = -dt^2 + a(t)^2 \left(dx_i^2 + K \frac{(x_i dx_i)^2}{1 - K x_i^2} \right) . \tag{1.15}$$

This is the so called Friedmann-Robertson-Walker (FRW) metric that is used to describe our universe. Note, that in addition to adding the time coordinate t, we have also allowed the scale factor $a(t)$ to change with time. This scale factor is the function that determines the evolution of our universe. In order to make this more transparent let us first go to spherical polar coordinates

$$dx_i^2 = dr^2 + r^2 d\Omega^2 , \qquad d\Omega^2 = d\theta^2 + \sin(\theta)^2 d\phi^2 , \qquad x_i dx_i = r dr , \tag{1.16}$$

so that the metric becomes

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dr^2}{1 - K r^2} + r^2 d\Omega^2 \right) . \tag{1.17}$$

Now we calculate the distance between an observer at the origin and an object at co-moving radial coordinate r (we always take $a(t) > 0$)

$$d(r,t) = a(t) \int_0^r \frac{dr'}{\sqrt{1-Kr'^2}} = a(t) \times \begin{cases} \arcsin(r) & K = +1 \\ \text{arcsinh}(r) & K = -1 \\ r & K = 0 \end{cases} . \tag{1.18}$$

This implies that any object at a fixed r moves away from us, if the scale factor $a(t)$ increases with time. More concretely, by differentiating the above equation we can establish the linear relationship between the distance and the velocity

$$\boxed{v = \frac{\partial d(r,t)}{\partial t} = \frac{\dot{a}(t)}{a(t)} d(r,t) \equiv H\, d(r,t)\,.} \tag{1.19}$$

In the last equation we defined

$$\boxed{H(t) = \frac{\dot{a}(t)}{a(t)}\,,} \tag{1.20}$$

where $H(t)$ is the so called *Hubble parameter* since Hubble discovered this linear relationship. So, we see that our FRW metric is correctly capturing Hubble's original observation in figure 3 provided that $\dot{a}(t) > 0$.

1.2.1 Interlude: setting $c = 1$

You are probably familiar with the fact that people use different units around the world. That is no problem since there are fixed conversion factors that convert one unit into another. The speed of light in vacuum is a similar universal conversion factor that we can use to convert time intervals into length and vice versa. This might seem peculiar to a person that has not studied special relativity. However, in special relativity you should have learned that space and time can mix under Lorentz transformation. So, they are not separate entities but rather unified into what is called spacetime.

Worked problem 1.4: Using different units

You are working on a ship and are in charge of calculating the distance it travels in spacetime. The crew tells you the travel time in minutes, the distance along the x_1-axis in nautical miles, the distance along the x_2-axis in yards and the distance along the x_3-axis in feet. Write down the corresponding line element ds^2 in flat spacetime (so-called Minkowski space) that gives ds in the SI unit of meters.

Solution: Similarly to above, we have to write down a line element with prefactors c and c_i, $i = 1, 2, 3$, that take care of the conversion from one unit to the other:

$$ds^2 = -c^2 dt^2 + c_1^2 dx_1^2 + c_2^2 dx_2^2 + c_3^2 dx_3^2\,. \tag{1.21}$$

The values of c and c_i can be determined to be

$$c = 1.8 \times 10^{10} \frac{m}{\text{minutes}}\,, \qquad c_1 = 1852 \frac{m}{\text{nautical miles}}\,,$$

$$c_2 = .91\frac{m}{\text{yard}}\,, \qquad c_3 = .3\frac{m}{\text{feet}}\,. \tag{1.22}$$

The problem above shows that we can use different units for different directions along spacetime. However, it also shows that this is very confusing. That is why we normally never do that and set $c = c_i = 1$. This does not fix the particular units we have to work with. However, it requires us to use the *same* units for all spacetime directions t, x_1, x_2, x_3. For example, if we decide that we work with meters m, then a light beam from the sun will have a travel *time* of roughly $\Delta t = 8 \cdot 1.8 \times 10^{10} m = 1.44 \times 10^{11} m$ to reach earth. On the other hand, if we decide that we work with minutes, then the *distance* from the earth to the sun is roughly $d = 8$ minutes.

1.3 The Friedmann equations

The evolution of the scale factor $a(t)$ is determined by the matter and energy content of the universe using general relativity. If you are not familiar with general relativity then the set of equations

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G\, T_{\mu\nu}\,, \tag{1.23}$$

might look rather intimidating and this paragraph might seem rather complicated. In this case you can jump ahead to the next paragraph. In an isotropic and homogeneous universe Einstein's equations boils down to two rather simple equations for $a(t)$. The left-hand-side of the above equation is entirely determined by the FRW metric given above and the cosmological constant Λ. The right-hand-side is determined by the energy-momentum tensor $T_{\mu\nu}$ that encodes the matter and energy in our universe. In a homogeneous and isotropic universe its spacial part has to be proportional to the metric $T_{ij} = p(t)g_{ij}$, where we allowed for an arbitrary time dependent function $p(t)$. The time component $T_{tt} = \rho(t)$ is also an arbitrary function. Finally, the mixed space-time components are a 3-vector. However, such a vector, if non-vanishing, would single out a particular direction which is inconsistent with isotropy that demands that the universe is the same in all directions so we have $T_{ti} = 0$.

Solving Einstein's equations above leads to the following two equations[2] that are called Friedmann's equations

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 + \frac{K}{a(t)^2} - \frac{\Lambda}{3} = \frac{8\pi G}{3}\rho(t)\,, \tag{1.24}$$

$$\frac{\ddot{a}(t)}{a(t)} - \frac{\Lambda}{3} = -\frac{4\pi G}{3}(\rho(t) + 3p(t))\,. \tag{1.25}$$

These equations involve three new quantities that deserve further discussion: The parameter Λ is called a cosmological constant and as we will see shortly, we can remove it from the equations by shifting ρ and p. So, this means that in a homogeneous

[2]Depending on your level of familiarity with general relativity I encourage you to either derive these equations yourself or to take a look at appendix A that gives the detailed derivation.

and isotropic universe we can describe any kind of matter, radiation and energy with just two quantities. What are these and how do we understand them intuitively? A homogeneous universe obviously requires a distribution of energy and matter that does not depend on the spacial coordinates, so instead of dealing with for example empty space dotted with galaxies we can take a continuum limit and think of it as a continuous distribution of matter. You might be familiar with similar approximations when describing air or water. Instead of describing all individual molecules, we describe the whole system as a continuous fluid. The quantity $\rho(t)$ describes the energy density (recall that mass equals energy due to $E = mc^2 = m$) and the function $p(t)$ describes the pressure of this fluid.

By looking at the equations (1.24), (1.25) we note that ρ and p can describe a cosmological constant. In particular, if we shift them such that

$$\rho \to \rho - \frac{\Lambda}{8\pi G}\,, \qquad p \to p + \frac{\Lambda}{8\pi G}\,, \tag{1.26}$$

then we remove Λ and find the Friedmann equations

$$\boxed{\begin{aligned} \left(\frac{\dot{a}(t)}{a(t)}\right)^2 + \frac{K}{a(t)^2} &= \frac{8\pi G}{3}\rho(t)\,, && (1.27) \\ \frac{\ddot{a}(t)}{a(t)} &= -\frac{4\pi G}{3}(\rho(t) + 3p(t))\,. && (1.28) \end{aligned}}$$

Worked problem 1.5: Absorbing K in ρ and p

Show that one can likewise absorb the curvature term $K/a(t)^2$ by shifting ρ and p.

Solution: By simple inspection we find that the following shifts do the trick

$$\rho \to \rho + \frac{3K}{8\pi G a(t)^2}\,, \qquad p \to p - \frac{K}{8\pi G a(t)^2}\,. \tag{1.29}$$

This would remove the $K/a(t)^2$ term and further simplify the Friedmann equations to

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3}\rho(t)\,, \tag{1.30}$$

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi G}{3}(\rho(t) + 3p(t))\,. \tag{1.31}$$

Given the above observation we are faced with the question of how to write the Friedmann equations and which terms should appear on the left-hand-side and which terms should be part of the right-hand-side. While this is somewhat of an academic question there is a reason of choosing equations (1.27) and (1.28) as our favorite choice. The way the Friedmann equations are derived leads naturally to the original form given in (1.24) and (1.25). These equations have a clear physical interpretation: The left-hand-side terms arise from general relativity and the right-hand-side arises from things that couple to gravity, like for example matter (electrons, protons, stars, etc.) or radiation (i.e., photons). So, in a hypothetical universe with only gravity and nothing else, the right-hand-side in equations (1.24) and (1.25) would be zero and the left-hand-side would be unchanged. However, as we will see towards the end of this course, there are particles that contribute to the right-hand-side of the equations (1.24) and (1.25) like a cosmological constant.[3] Therefore, it makes sense to remove Λ on the left-hand-side and include it into ρ and p on the right-hand-side, i.e., to work with the Friedmann equations as given in equations (1.27) and (1.28). No particles are known that would contribute to the right-hand-side like the curvature term $K/a(t)^2$. Therefore, we keep the curvature term on the left-hand-side.

The two rather simple Friedmann equations govern our universe from a split second after the big bang until today. All we need for this is the knowledge of $\rho(t)$ and $p(t)$, i.e., of the matter and energy content of our universe. As we will see in the next lecture, these functions are not too complicated and usually at each time there is one form of energy that is dominating the expressions so that we can solve the Friedmann equations analytically.

Worked problem 1.6: Constraints on the simplest solutions

Show that in a universe with only gravity, i.e., $\rho = p = 0$ in equations (1.24) and (1.25), not all values of K are compatible with $\Lambda \leq 0$.

Solution: The first Friedmann equation simplifies to

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 + \frac{K}{a(t)^2} - \frac{\Lambda}{3} = 0\,. \tag{1.32}$$

If $\Lambda < 0$ then the left-hand-side is strictly positive for $K = 0$ and $K = +1$ and we have a contradiction. So, only $K = -1$ is allowed.
If $\Lambda = 0$, then $K = 0$ and $K = -1$ are allowed but $K = +1$ is again forbidden.

Differentiating (1.27) we get

$$\frac{8\pi G}{3}\dot{\rho}(t) = 2\frac{\dot{a}(t)}{a(t)}\left(\frac{\ddot{a}(t)}{a(t)} - \left(\frac{\dot{a}(t)}{a(t)}\right)^2 - \frac{K}{a(t)^2}\right)\,. \tag{1.33}$$

Using now equation (1.27) and (1.28) and recalling that $H = \dot{a}/a$ we find the *continuity equation*:

$$\boxed{\dot{\rho}(t) + 3H(t)\left(\rho(t) + p(t)\right) = 0\,.} \tag{1.34}$$

This equation will be useful, when we discuss the different matter and energy content of the universe in the next section.

[3] From quantum gravity considerations it is also not clear whether a constant Λ can exist.

Summary: Friedman equations

There are two long range forces in the universe, electromagnetism and gravity. Since the universe is electrically neutral on large scales, gravity dominates and the evolution of our universe is described entirely by the theory of general relativity.
Our universe is, on sufficiently large scales of several Mpc, homogeneous and isotropic. This dramatically simplifies the equation of motions for general relativity to the two Friedman equations:

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 + \frac{K}{a(t)^2} = \frac{8\pi G}{3}\rho(t)\,, \tag{1.35}$$

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi G}{3}\left(\rho(t) + 3p(t)\right). \tag{1.36}$$

The two equation above can be combined into (and any one of them can be replaced with) the continuity equation:

$$\dot{\rho}(t) + 3H(t)\left(\rho(t) + p(t)\right) = 0\,. \tag{1.37}$$

In the equations above $a(t)$ is the time dependent scale factor that controls the size of spatial parts of our universe, $K = 0, \pm 1$ is the curvature, $G = 6.67 \times 10^{-11} m^3 kg^{-1} s^{-2}$ is Newton's constant and $\rho(t)$ and $p(t)$ are the energy density and pressure of the homogeneous fluids that fill our universe. We also defined the Hubble parameter as

$$H(t) = \frac{\dot{a}(t)}{a(t)}\,. \tag{1.38}$$

2 Dynamics of the universe

In the last section we learned that our universe is homogeneous and isotropic and can be described by the so called FRW metric. Using general relativity one can derive the Friedmann equations that describe the evolution of the universe for any given energy and matter content with an energy density given by $\rho(t)$ and a pressure $p(t)$. In this section we will discuss different forms of matter, work through Einstein's biggest blunder and determine the age and time evolution of our universe.

2.1 The different forms of matter

There are three different forms of matter and energy in our universe and they all satisfy the relation $p(t) = w\rho(t)$, where the constant w is called the equation of state parameter. Plugging this into the continuity equation (1.34) we can derive the following

$$0 = \dot{\rho}(t) + 3\frac{\dot{a}(t)}{a(t)}(1+w)\rho(t)$$

$$
\begin{aligned}
0 &= \frac{\dot{\rho}(t)}{\rho(t)} + 3(1+w)\frac{\dot{a}(t)}{a(t)} \\
0 &= \frac{d}{dt}\ln\rho(t) + 3(1+w)\frac{d}{dt}\ln a(t) \\
0 &= \ln(\rho(t)) + 3(1+w)\ln(a(t)) + const. \\
0 &= \ln(\rho(t)) + \ln\left(a(t)^{3(1+w)}\right) + const. \\
1 &= \rho(t)\cdot a(t)^{3(1+w)}\cdot e^{const.} \\
&\Rightarrow \quad \rho(t) \propto a(t)^{-3(1+w)}\,. \qquad (2.1)
\end{aligned}
$$

Thus, we see that the function $\rho(t)$ and therefore $p(t) = w\rho(t)$ are related to $a(t)$ in a rather simple way. This means, as we will see below, that as long as a single matter component is dominating, we can solve the equations that determine the evolution of our entire universe analytically!

Worked problem 2.1: The initial big bang singularity

Use the above derived equation (2.1) to discuss what happens to the energy density in the very early universe. In particular what happens to a pressureless fluid with $p = 0$?

Solution: From Hubble's observation we know that our universe is expanding. That means that it was smaller in the past and in particular much smaller in the early universe. So, since $a(t)$ decreases, $\rho(t)$ will likewise change, unless $w = -1$. For $w < -1$ we find that ρ goes to zero. For the case of a pressureless fluid we have $p = w\rho = 0$, so, $w = 0$ (since the energy density is positive). For such $w = 0$ and more generally for all $w > -1$ we find that the energy density goes to infinity when $a(t)$ goes to zero.

Before we solve the Friedmann equations let us discuss what kind of matter and energy we expect to have in our universe and derive the corresponding equation of state parameter w.

- **Non-relativistic matter**

  The matter we are most familiar with are stars and galaxies that we can observe at night in the sky. This form of matter has a velocity that is much smaller than the speed of light so that we can neglect its kinetic energy. In a given box in which each side has the initial length $a(t_{in})l$, we have a certain number of stars and galaxies with a mass M. The energy density is then given by $\rho = E/(a(t_{in})l)^3 = M/(a(t_{in})l)^3$, where we used $E = M$ in units where $c = 1$. Now when the universe evolves, the box will change its volume to $a(t)^3 l^3$ as is shown in figure 9.

  Since the mass M stays the same we find the following scaling

$$
\boxed{\rho_m(t) \propto a(t)^{-3} \quad \Leftrightarrow \quad w = 0 \quad \Leftrightarrow \quad p_m(t) = 0\,.} \qquad (2.2)
$$

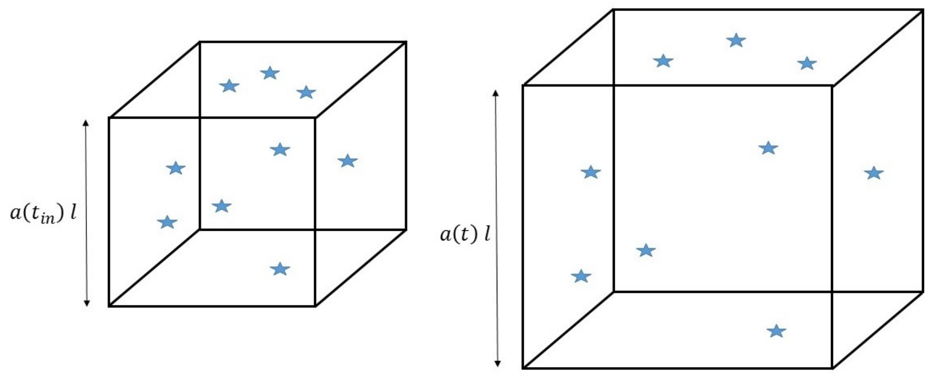


Figure 9: Non-relativistic matter in an expanding universe.

So, we see that non-relativistic matter has an equation of state parameter $w = 0$ and therefore vanishing pressure, which makes sense since the matter inside our box should not exert any pressure on the walls.

As we will discuss below, the largest fraction of cold (i.e., non-relativistic) matter in our universe is in the form of an unknown so called *dark matter*.

- **Radiation**

  Another form of energy in the universe is radiation (like for example light). The energy of light in units where $c = \hbar = 1$ is given by $E = 2\pi/(a(t_{in})\lambda)$, where $a(t_{in})\lambda$ is the wavelength. If we have a certain number of photons inside a big volume of initial size $(a(t_{in})l)^3$, then the energy density gets again diluted due to the increase in the volume of the box as above. Additionally, due to the expansion of the space the initial wavelength $a(t_{in})\lambda$ increase to $a(t)\lambda$, as shown in figure 10,

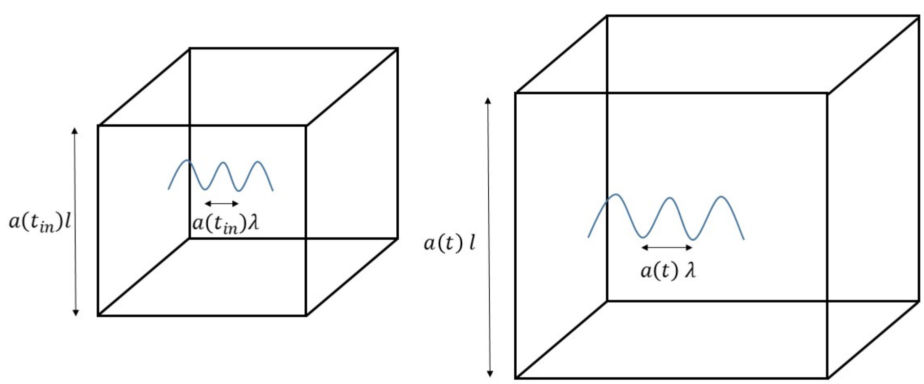


Figure 10: Radiation in an expanding universe.

so that we find for radiation

$$\boxed{\rho_r(t) \propto a(t)^{-4} \quad \Leftrightarrow \quad w = \frac{1}{3} \quad \Leftrightarrow \quad p_r(t) = \frac{1}{3}\rho_r(t)\,.} \tag{2.3}$$

As we will learn later, our universe is filled with the cosmic microwave background, which is thermal radiation left over from the big bang. Its spectrum is the best measured black body in nature.

- **The cosmological constant**

  As we have seen in subsection 1.3 in equations (1.24), (1.25) and(1.26), we can describe a cosmological constant by

$$\boxed{\rho_\Lambda(t) = -p_\Lambda(t) = \frac{\Lambda}{8\pi G} \quad \Leftrightarrow \quad w = -1\,.} \tag{2.4}$$

  So, in this case $\rho_\Lambda = -p_\Lambda$ is constant and the energy density does not change in time. This can be understood as follows: During the expansion of the universe more of the vacuum is created and this vacuum has a non-zero energy density ρ so that ρ does not change during the expansion (or contraction), as shown in figure 11.

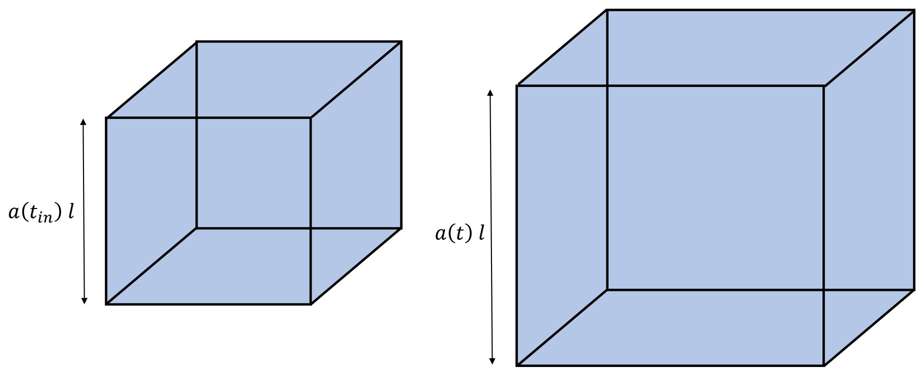


Figure 11: A cosmological constant does not get diluted in an expanding universe.

2.2 Conservation of energy

If you are not too familiar with the theory of general relativity, then you might wonder about the conservation of energy in the above examples. This is a generic feature of general relativity. The would-be conservation of energy is replaced by the condition that $\nabla_\mu T^{\mu\nu} = 0$, where ∇_μ denotes the covariant derivative. In particular that means that $\partial_\mu T^{\mu\nu} + \Gamma^\nu_{\mu\sigma} T^{\mu\sigma} + \Gamma^\mu_{\mu\sigma} T^{\sigma\nu} = 0$. Using appendix A that derives Friedmann's equations you can check that the above four equations ($\nu = 0, 1, 2, 3$) reduce to the continuity equation (1.34) for $\nu = 0$ and are trivial otherwise. If you are confused about how this non-conservation of energy is possible in a physical theory, recall that the conservation of energy follows via Noether's theorem from the time-translational symmetry. So, any physical theory that is not invariant under time translations can and generically will violate the conservation of energy. An expanding universe is certainly not invariant under time translations so it does violate the standard conservation of energy but it does satisfy the continuity equation that was implied by the two Friedmann equations.

Worked problem 2.2: Energy conservation in a static universe

While an expanding or contracting universe is generically not invariant under time translations, a static universe is. Show that energy is conserved in a static universe.

Solution: We have learned above, that energy conservation is generalized to the continuity equation (1.34):

$$\dot{\rho}(t) + 3H(t)\left(\rho(t) + p(t)\right) = 0\,. \tag{2.5}$$

In a static universe $\dot{a} = 0$ and therefore $H = \dot{a}/a = 0$ and the continuity equation reduces to $\dot{\rho} = 0$. Since the universe is static, volumes do not change and we can multiply ρ by an arbitrary volume V and find that the energy in that volume is conserved

$$\dot{E} = \frac{d(\rho V)}{dt} = \dot{\rho} V = 0\,. \tag{2.6}$$

2.3 The dust filled universe

In this section we will study the simple case of a universe which contains non-relativistic matter, so we set $p(t) = 0$ and we have $\rho(t) \propto a(t)^{-3} > 0$ from (2.2). The second Friedmann equation (1.28) then immediately tells us that such a universe cannot be static, i.e., $\ddot{a}(t) \neq 0$. In fact, it tells us that the expansion of the universe is decelerating. This is very intuitive since we know that gravity always attracts. In a universe filled with matter the gravitational attraction between the matter will slow down any initial expansion. There then seem to be three possibilities:

1. The universe will keep expanding forever at a slower and slower rate.
2. The expansion of the universe will eventually come to a stop.
3. The expansion will slow down and then gravitational attraction between the matter forces the universe to contract and eventually collapse.

We will see that these cases correspond to $K = -1, 0, 1$. We can write equation (1.27) as

$$0 \leq \dot{a}(t)^2 = \frac{8\pi G}{3}\rho_m(t)a(t)^2 - K = \frac{c_m}{a(t)} - K\,, \tag{2.7}$$

where we introduced the constant $c_m > 0$ via $8\pi G\,\rho_m(t)/3 = c_m/a(t)^3$ and used (2.2). We immediately see that for $K = -1$ the right-hand-side can never vanish so in this universe any initial expansion $\dot{a}(t)$ will go on forever. For the case $K = 0$ the right-hand-side vanishes for $a(t) \to \infty$ so the expansion will eventually come to a stop. Finally, for $K = +1$ the first term dominates for very small $a(t)$ but once $a(t) = c_m/K$ the expansion will come to a stop and the universe will then contract (since (1.27) implies $\ddot{a}(t) < 0$). These three scenarios are shown in figure 12.

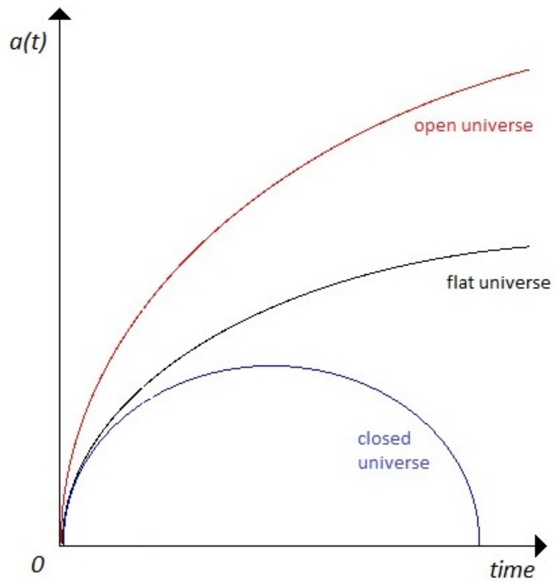


Figure 12: An open, flat and closed universe corresponding to $K = -1, 0, 1$.

We can also rewrite the first Friedmann equation as given in (2.7) as

$$\frac{1}{2}\dot{a}(t)^2 + V(a(t)) = -\frac{K}{2}\,, \tag{2.8}$$

with $V(a(t)) = -c_m/(2a(t))$. The above equation describes the motion of a 1-dimensional particle in the potential V and with a total energy $E = -K/2$. Since $V(a(t)) < 0$ we can conclude that for $E \geq 0$, i.e., $K = 0$ and $K = -1$ there exist unbound solutions and the universe can expand forever. For $K = 1$ we have $E = -1/2$ and the trajectories are bound. This is shown in figure 13.

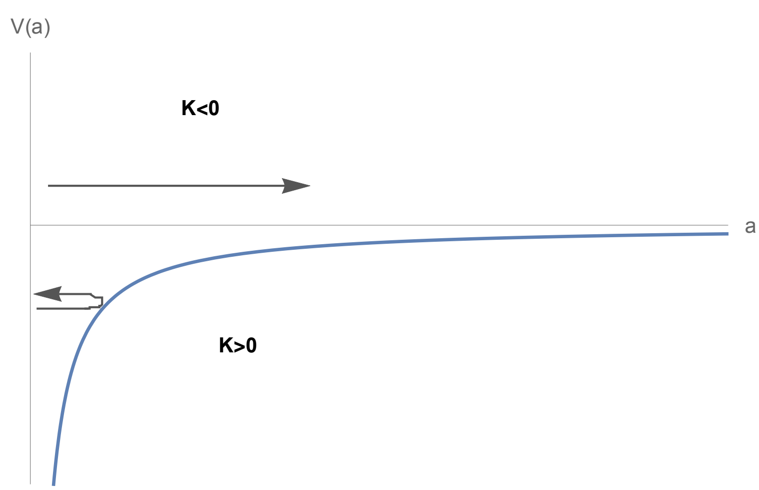


Figure 13: Unbound trajectories only exist for $K = -1$ and $K = 0$, while for $K = 1$ the universe expands and then contracts again.

2.3.1 A static universe?

Note, that independent of the value of K we find that the second Friedmann equation (1.28) does not permit static solutions for a universe filled with matter. Before the discovery that our universe was expanding, this fact was very troublesome for people like Einstein that imagined our universe to be time independent. Let us therefore try to construct a static universe with matter by adding in the cosmological constant Λ. The universe can then be described by $\rho(t) = \rho_m(t) + \rho_\Lambda$, $p = p_\Lambda = -\rho_\Lambda$. In a static universe with $\dot{a}(t) = \ddot{a}(t) = 0$, we then find from equation (1.28) that

$$0 = \rho + 3p = \rho_m - 2\rho_\Lambda \quad \Leftrightarrow \quad \rho_m = 2\rho_\Lambda > 0\,, \tag{2.9}$$

since $\rho_m > 0$. Using this in equation (1.27) gives

$$\frac{K}{a^2} = \frac{8\pi G}{3}\rho = \frac{8\pi G}{3}(\rho_m + \rho_\Lambda) = 8\pi G \rho_\Lambda = \Lambda > 0\,. \tag{2.10}$$

So, we have succeeded in finding a static solution provided that $K = 1$ and $\Lambda > 0$. In this static solution we have $a = 1/\sqrt{\Lambda}$. The important question to ask is whether such a solution is stable. To answer that, we can look again at equation (2.8). The potential now has an extra contribution from the cosmological constant so that we find for $K = 1$

$$\frac{1}{2}\dot{a}(t)^2 + V(a(t)) = \frac{1}{2}\dot{a}(t)^2 - \frac{c_m}{2\,a(t)} - \frac{1}{6}\Lambda\; a(t)^2 = -\frac{1}{2}\,. \tag{2.11}$$

A plot of the potential is shown in figure 14.

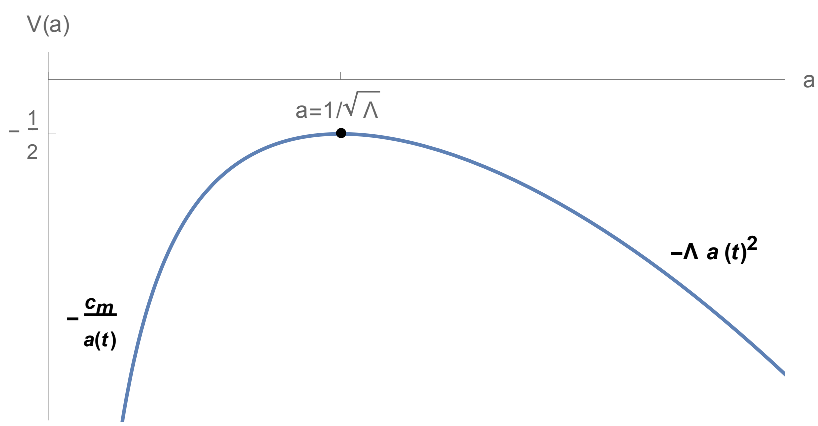


Figure 14: The potential for a static universe with $K = +1$, matter and a cosmological constant $\Lambda > 0$. We see that the static solution with $a = 1/\sqrt{\Lambda}$ is a maximum and therefore unstable.

We see that our static universe corresponds to a maximum of the potential. This means this static universe is unstable. If the matter and energy content is the tiniest bit different, then this universe will either expand forever or collapse. Fortunately, our universe is not static, so that we don't have to worry about such delicate solutions.

Historically, the following happened: In 1915 Einstein's published his theory of general relativity without the cosmological constant Λ. When he then tried to use his equations to describe our universe, which he believed to be static, he failed. Instead of trusting his theory and predicting that our universe is not static, he introduced the cosmological constant Λ and gives it the very specific value required for the static universe discussed in this subsection. Later Einstein will call this his ‘biggest blunder’. In 1927 Georges Lemaître shows, by combining general relativity with Hubble’s observations, that our universe is expanding. In 1931 Einstein finally accepts the idea of an expanding universe and he proposes together with Willem de Sitter the model of an expanding universe without a cosmological constant, i.e., the model that we discussed at the beginning of subsection 2.3.

2.3.2 The age of the universe

Since we know that our universe is expanding, let us ask the simple question of how old our universe would be, if all its energy would be contained in non-relativistic matter. This is not that bad of an approximation and will give us an age that is of the correct order of magnitude. Before we start the calculation let us introduce an important convention. We call our current time t_0, i.e., we use a subscript 0 to denote today’s value of the time variable. Likewise, we use $a_0 = a(t_0)$ and $H_0 = H(t_0) = \dot{a}(t_0)/a(t_0)$ to denote today’s value of the scale factor and Hubble parameter. Since H_0 by definition is a constant, it is usually called the Hubble constant. There are ever improving measurements of the Hubble constant but its uncertainty is still somewhat large. For that reason, one usually writes

$$H_0 = 100h\,\frac{km}{s\,Mpc}\,, \tag{2.12}$$

where the current experimental value of h is[4]

$$h = .677 \pm .008\,. \tag{2.13}$$

Hubble’s original observations led him to $h \approx 5$ due to several systematic errors, cf. figure 3. So, over the last century astrophysicists reduced the error from a few hundred percent to just a few percent.

The unit of H and H_0 is 1/time. This can be easily seen from the definition $H = \dot{a}/a$, where the length dimension of a cancels but the time derivative leaves us with 1/time. The reason why there are two different length units in the quoted value for H_0 in equation (2.12) is that astronomers measure cosmological distances in *Mpc* and velocities in km/s. So, looking back at Hubble’s law in equation (1.19): $v = H\ d$, we see why the unusual units arise.

Worked problem 2.3: The inverse Hubble constant in years

What is $1/H_0$ in years?

[4] As we will discuss later, there are two very different ways of determining the Hubble constant and they currently disagree with each other. It is not clear right now why that is the case. The value given below is obtained by the Planck collaboration by studying the cosmic microwave background.

Solution: We simply need to convert Mpc to km and seconds to years

$$\frac{1}{H_0} = \frac{s\,Mpc}{67.7km} = \frac{yr}{3.15\times 10^7 s}\frac{s\ 3.09\times 10^{19}km}{67.7km} = 14.4\times 10^9 yr\,. \quad (2.14)$$

This value of 14.4 billion years is actually fairly close to the age of our universe. As we will see below, this is not a coincidence.

Now let us use the value of H_0 to determine the age of a universe filled with non-relativistic matter. We will set $K = 0$ which, as we will discuss below, is very much consistent with observation. We then find from equation (2.2) that

$$\rho(t) = \rho_0 \left(\frac{a_0}{a(t)}\right)^3 , \quad (2.15)$$

where ρ_0 is the current energy density of the universe. Now we use this in the Friedmann equation (1.27)

$$\begin{aligned} a(t)\dot{a}(t)^2 &= \frac{8\pi G}{3}\rho_0 a_0^3 \\ \sqrt{a(t)}\ \dot{a}(t) &= \sqrt{\frac{8\pi G}{3}\rho_0 a_0^3} \\ \sqrt{a(t)}\ da &= \sqrt{\frac{8\pi G}{3}\rho_0 a_0^3}\ dt\,, \end{aligned} \quad (2.16)$$

where in the last line we used $\dot{a}(t) = da/dt$. Now we can integrate both sides which leads to

$$\frac{2}{3}a(t)^{\frac{3}{2}} = \sqrt{\frac{8\pi G}{3}\rho_0 a_0^3}\ t + const. \quad (2.17)$$

By demanding that $a(t = 0) = 0$ is the initial singularity we find that the integration constant vanishes. The above equation implies

$$a(t) = a_0 \left(\frac{t}{t_0}\right)^{2/3} , \quad (2.18)$$

since by definition $a(t_0) = a_0$ and t_0 is implicitly defined in (2.17) but we will not need this particular expression. We rather calculate H_0 directly

$$H_0 = \frac{\dot{a}(t_0)}{a(t_0)} = \frac{\frac{2}{3}a_0\, t_0^{-\frac{1}{3}}/t_0^{\frac{2}{3}}}{a_0} = \frac{2}{3}\frac{1}{t_0}\,. \quad (2.19)$$

So, we have found the age of a matter filled universe in terms of the Hubble constant today

$$t_0 = \frac{2}{3}\frac{1}{H_0} = \frac{2}{3}\frac{s\,Mpc}{100h\,km} \approx \frac{2}{300h}3\times 10^{19}\ s \approx 9.6\times 10^9\ yr = 9.6\ Gyr\,. \quad (2.20)$$

While this is pretty close to the age of our universe which is roughly 13.8×10^9 years, it is inconsistent with the observation of the oldest stars that are as old as 13×10^9 years.

2.4 Time evolution of the universe

As we have seen above, a matter dominated universe gives us the right order of magnitude for the age of the universe but the answer is inconsistent with observations. The reason for that is that our universe contains other forms of energy as well. At different points of time different forms of energy density dominate the evolution of the universe. Let us therefore also determine the time dependence of the scale factor for the other cases. We start with the general expression (2.1) which is equal to

$$\rho(t) = \rho_0 \left(\frac{a(t)}{a_0}\right)^{-3(1+w)} . \tag{2.21}$$

Using this in the Friedmann equation (1.27) and assuming a negligible curvature contribution ($K = 0$), we can repeat the above calculation

$$\begin{aligned} \left(\frac{\dot{a}(t)}{a(t)}\right)^2 &= \frac{8\pi G}{3}\rho_0 \left(\frac{a(t)}{a_0}\right)^{-3(1+w)} \\ a(t)^{\frac{1+3w}{2}}\dot{a}(t) &= \sqrt{\frac{8\pi G}{3}\rho_0 a_0^{3(1+w)}} \\ a(t)^{\frac{1+3w}{2}} da &= \sqrt{\frac{8\pi G}{3}\rho_0 a_0^{3(1+w)}}\, dt \\ \frac{2}{3(1+w)} a^{\frac{3(1+w)}{2}} &= \sqrt{\frac{8\pi G}{3}\rho_0 a_0^{3(1+w)}}\, t + const. , \qquad w \neq -1 . \end{aligned} \tag{2.22}$$

We can again set the constant to zero by choosing $a(t = 0) = 0$ and fix the factor of proportionality by demanding that $a(t_0) = a_0$ and get

$$a(t) = a_0 \left(\frac{t}{t_0}\right)^{\frac{2}{3(1+w)}} , \quad w \neq -1 . \tag{2.23}$$

The above derivation doesn't apply to the case of a cosmological constant but in that case one has simply $\rho(t) = const.$ and finds from (1.27) that

$$a(t) = a_0 e^{H_0(t-t_0)} . \tag{2.24}$$

Note that in this case the 'beginning' of the universe is not at $t = 0$ but rather at $t = -\infty$. So, such a universe is infinitely old. This case is also special since the Hubble parameter $H(t)$ is actually constant (since ρ is constant), while in all other cases it changes with time as

$$H(t) = \frac{2}{3(1+w)\,t} , \quad w \neq -1 . \tag{2.25}$$

Our derivation above also applies to the case of a negatively curved universe with $K = -1$ and $\rho(t) = 0$, since this can be thought of as a fluid with energy density $\rho \propto a(t)^{-2}$ which is equal to $w = -1/3$. Let us summarize the different scalings we found

$$
\begin{aligned}
a(t) =& a_0 \left(\frac{t}{t_0}\right)^{\frac{1}{2}}, && \text{for radiation, i.e., } w = \frac{1}{3}, && (2.26)\\
a(t) =& a_0 \left(\frac{t}{t_0}\right)^{\frac{2}{3}}, && \text{for matter, i.e., } w = 0, && (2.27)\\
a(t) =& a_0 \frac{t}{t_0}, && \text{for curvature with } K = -1 \text{, i.e., } w = -\frac{1}{3}, && (2.28)\\
a(t) =& a_0 e^{H_0(t-t_0)}, && \text{for } \Lambda \text{, i.e., } w = -1. && (2.29)
\end{aligned}
$$

Worked problem 2.4: Energy density and pressure

Determine the time dependence of the energy density ρ and the pressure p.

Solution: For a cosmological constant with $w = -1$ we have that $\rho = -p = const$. So, they both do not change in time.
Otherwise, we find for $\rho(t)$ (and therefore also $p(t) = w\rho(t)$) from the first Friedmann equation (1.27) and using equation (2.25) that

$$\rho(t) = \frac{3}{8\pi G} H(t)^2 = \frac{3}{8\pi G} \left(\frac{2}{3(1+w)}\right)^2 \frac{1}{t^2}, \quad w \neq -1. \tag{2.30}$$

This means that, in the absence of a cosmological constant, in an expanding universe the energy density is diluted as t^{-2} independently of which kind of fluid dominates the energy density.

2.5 Fun facts

Let us conclude with two non-trivial observation. A universe with a flat geometry $K = 0$ is special in the sense that it can lead to a critical evolution (see figure 12) but it also is special since our own universe seems to have a very small (or even vanishing) curvature $|K/a_0^2| \ll (\dot{a}(t_0)/a_0)^2$. We have seen in the first lecture that the spatial part of such a universe could be simple the flat space $\mathbb{R}^3$. This is however not the entire truth. It is also possible that one, two or all of the three x_i directions are periodic, i.e., they are circles. This would mean, if these circles wouldn't be too large, we could see ourselves in the sky or we could see the same galaxy twice in the universe by looking in opposite directions. However, up to date there is no evidence of such a non-trivial topology so if the spacial part of our universe is finite (or periodic in any one direction), then the corresponding radius has to be very large and we might never be able to observe this. However, it is interesting to know that our universe (or more precisely a universe with $K = 0$) does not necessarily have to be spatially infinite.

A common question that arises when discussing the big bang is its location. Where did it happen, i.e., where is the center of the universe? We have learned that our universe is homogeneous and isotropic. This means there is no special point at which the big bang could have happened. Rather it must have happened everywhere. We can understand this intuitively from the FRW metric (see equation (1.15))

$$ds^2 = -dt^2 + a(t)^2 \left(dx_i^2 + K \frac{(x_i dx_i)^2}{1 - K x_i^2} \right) . \tag{2.31}$$

The big bang singularity is in principle the point in the past where $a(t) = 0$. At this point $dx_i^2 + K(x_i dx_i)^2/(1 - K x_i^2)$ still describes flat space, a 3-sphere or a hyperspherical surface with negative curvature, however, this part gets multiplied by zero. For a 3-sphere, which has a finite coordinates range (recall $x_i^2 \leq 1$) and for flat space with three periodic directions, this means that the corresponding compact spaces shrink to zero size at the big bang singularity. However, for $\mathbb{R}^3$ and the hyperspherical surface the coordinate ranges are infinite. So, in this case the universe remains infinitely large even at the big bang singularity! Why is it then called a singularity if the universe might have still been infinitely large? The answer is for example given in problem 2.4 above. At the big bang singularity certain quantifies diverge and our equations break down. Strictly speaking we should think of general relativity as a 'low energy effective theory' that is only valid for energies below the Planck energy. In particular, we should only use it at times $t > t_P = 5.4 \times 10^{-44} s$.

Summary: The dynamics of the universe

In this section we learned about the time evolution of our universe that is determined by the fluid that dominates the energy density of the universe. Three such fluids are non-relativistic matter like stars and galaxies, radiation and dark energy. All of these can be described by fluids with $p(t) = w\rho(t)$, where the constant w is called the equation of state parameter. It takes the vales $w = 0$ for matter, $w = 1/3$ for radiation and $w = -1$ for a cosmological constant.

We then derived how the scale factor and the energy density change with time and found

$$a(t) = a_0 \left(\frac{t}{t_0} \right)^{\frac{2}{3(1+w)}} , \quad w \neq -1 , \tag{2.32}$$

and

$$\rho(t) = \frac{3}{8\pi G} \left(\frac{2}{3(1+w)} \right)^2 \frac{1}{t^2} , \quad w \neq -1 . \tag{2.33}$$

For the special case of a cosmological constant with $w = -1$ one finds that the energy density is constant and that

$$a(t) = a_0 e^{H_0 (t - t_0)} . \tag{2.34}$$

3 Our universe (and its fate)

In this section we discuss the observed values for the different forms of energy and matter in our universe. Based on that it is easy to discuss the fate of our universe in the far distant future. It will be however much more interesting and complicated to describe the history of our universe from the beginning until today and that is what is going to occupy us for the rest of this semester.

3.1 Critical density

In a universe like our own the curvature $|K/a_0^2| \ll (\dot{a}(t_0)/a_0)^2$ is very small and it is useful to define the so-called critical energy density. From the first Friedmann equation

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 + \frac{K}{a(t)^2} = \frac{8\pi G}{3}\rho(t)\,, \tag{3.1}$$

we find after setting $K = 0$ the critical energy density

$$\boxed{\rho_c(t) = \frac{3H(t)^2}{8\pi G}\,.} \tag{3.2}$$

This critical density today is roughly $\rho_c(t_0) \approx 10^{-26} kg/m^3$ which is incredibly small.

Worked problem 3.1: The best vacuum obtained on earth

The best vacuum obtainable on earth with our current technology has around 1,000 atoms/cm^3. Assume the atoms are the lightest possible atoms, namely hydrogen atoms. What is the energy density in such a vacuum?

Solution: 1,000 hydrogen atoms have a mass of $m = 1.7 \times 10^{-24} kg$. This leads to an energy density of

$$\rho_{vac} = 1.7 \times 10^{-24} kg/cm^3 = 1.7 \times 10^{-18} kg/m^3\,. \tag{3.3}$$

Interplanetary space contains roughly 11 molecules per cm^3, interstellar space 1 molecule and intergalactic space 10^{-6} molecules per cm^3.
Since there is a lot of intergalactic space in the universe, the critical energy density $\rho_c(t_0) \approx 10^{-26} kg/m^3$ above is not much larger than the energy density for intergalactic space $\rho_{int-gal} \approx 1.7 \times 10^{-27} kg/m^3$.

Having defined the critical density, we can normalize the energy density for all fluids by dividing by the critical density and define

$$\Omega_i(t) \equiv \frac{\rho_i(t)}{\rho_c(t)} \quad \text{with} \quad i = m, r, \Lambda\,. \tag{3.4}$$

We can also define the total normalized energy density $\Omega_{tot} = \sum_i \Omega_i$. With that we can rewrite equation (3.1) as

$$\Omega_{tot}(t) = \sum_i \Omega_i(t) = 1 + \frac{K}{\dot{a}(t)^2} \,. \tag{3.5}$$

From this we see that for an open universe $\Omega_{tot} < 1$ and for a closed universe $\Omega_{tot} > 1$. For a flat universe with $K = 0$ we have $\Omega_{tot} = 1$ being constant. Note, that observations will always have uncertainties. This means that if our universe is actually flat ($K = 0$) then we would never be able to know for sure. However, if our universe would have non zero curvature $K/\dot{a}(t)^2 \neq 0$, then this is something we could in principle measure with very high confidence.

3.2 Our universe

In the last few decades our understanding of the universe has substantially improved and cosmology has become a precision science where most parameters can be measured with error bars that are a few percent or even less. This 'golden age' of cosmology is far from over and future experiments promise substantially better measurements and hold the prospect of discovering new fascinating features of our universe.

As we have discussed before, $|K/\dot{a}_0^2| \ll 1$ in our current universe and there is no observational evidence for non-vanishing curvature. The current upper bound is

$$\left|\frac{K}{\dot{a}_0^2}\right| < .005 \,. \tag{3.6}$$

As discussed above, this means that the total energy density of our universe is very close to the critical energy density. This then for example means that if a particular type of matter or energy constitutes for example 70% of the energy density in our current universe, then its current energy density is $.7\rho_c(t_0)$.

Worked problem 3.2: The value of a_0 in the absence of curvature

Show that for $K = 0$ the Friedmann equations (1.27) and (1.28) are invariant under rescaling of $a(t)$ by an arbitrary constant. What does that mean for the value of a_0?

Solution: Rescaling $a(t)$ by a constant c we find for $K = 0$

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \left(\frac{c\,\dot{a}(t)}{c\,a(t)}\right)^2 = \frac{8\pi G}{3}\rho(t) \,, \tag{3.7}$$

$$\frac{\ddot{a}(t)}{a(t)} = \frac{c\,\ddot{a}(t)}{c\,a(t)} = -\frac{4\pi G}{3}(\rho(t) + 3p(t)) \,. \tag{3.8}$$

So, the Friedmann equations are indeed invariant under this rescaling. This means that the value of a_0 has no physical meaning and cancels out in all calculations for $K = 0$. We can understand this more mathematically by the fact that in flat space we can rescale the x_i-coordinates by $1/c$,

while rescaling $a(t)$ by c. This does not change the metric and leads to a physically equivalent situation.
Note, that this is independent of whether some or all of the x_i coordinates are compact or not.
Note also that this means that in our universe where $\left|\frac{K}{a_0^2}\right|$ is unmeasurably small, the value of a_0 is arbitrary and has nothing to do with the size of our universe.

Let us now take inventory of our current universe:

- **Matter:** While matter like the stars in the galaxies are the most obvious form of matter one can think of, they turn out to be actually only a minimal fraction of the matter in our universe (roughly 0.5%). However, there is a lot of Hydrogen and Helium in the universe in large clouds that contribute to what is usually called *baryonic matter*. The reason for this name is presumably that the baryons (protons and neutrons) make up for almost the entire amount of the mass and the leptons (electrons) only contribute a very small amount of mass. The current value for the baryonic density parameter is

$$\Omega_{b,0}h^2 = .02226 \pm .00023 \quad \Rightarrow \quad \Omega_{b,0} \approx .048\,. \tag{3.9}$$

  There is another form of matter that only recently became non-relativistic which are neutrinos. They contribute roughly 0.3% of the total energy density. As we will discuss later, during the early times of the universe neutrinos behaved as radiation and not as matter.

  So, the matter we know and understand constitutes only roughly 5% of the total energy density of our current universe!

  It turns out that there has to be another type of non-relativistic matter in our universe in order to explain for example the mass difference between the mass of the visible matter in galaxies and the total mass derived from gravitational effects. Since this matter is not visible because it doesn't interact with photons, it is called *dark matter*. We still don't know what this dark matter really is and have not yet been able to detect dark matter particles in any of the ongoing experiments. Nevertheless, we can conclude from cosmological observations that their contribution to the density parameter is

$$\Omega_{c,0}h^2 = .1186 \pm .0020 \quad \Rightarrow \quad \Omega_{c,0} \approx .258\,. \tag{3.10}$$

  So, the energy density of our current universe arises to roughly 30% from non-relativistic matter with an equation of state parameter $w = 0$. The exact current bound is

$$\boxed{\Omega_{m,0} = .308 \pm .012\,.} \tag{3.11}$$

- **Radiation:** We know that there are photons (light) in our universe, however, these contribute a negligible amount to the current energy density. In particular, the photons from the cosmic microwave background, that will play a very important role in the coming lectures, contribute to the density parameter only

$$\boxed{\Omega_{r,0} \approx 5 \times 10^{-5}\,.} \tag{3.12}$$

Our universe also contains gravitons and the corresponding gravitational waves that contribute to the radiation. Gravitational waves have first been detected directly in 2015 by the LIGO collaboration. The so far detected gravitational waves arise from mergers of very massive objects like neutron stars and black holes. There are also other sources of gravitational waves in our current universe and we expect that the universe should additionally be filled with a primordial cosmic gravitational wave background. However, the contributions from all these sources to the radiation density parameter today is negligible.

So, we find that radiation is unimportant in our current universe. However, due to its equation of state parameter $w = 1/3$ we derived $\Omega_r(t) \propto a(t)^{-4}$. This means that in the early universe where $a(t)$ was much smaller, radiation was actually the dominating form of energy.

- **Dark Energy:** The largest contribution to the density parameter in our current universe is due to *dark energy*, a currently not fully understood form of energy with negative pressure that leads to an accelerated expansion of our universe. We will discuss this in more detail in the next section. Here let us just say that dark energy is very compatible with a cosmological constant Λ with equation of state $w = -1$. The current contribution to the density parameter is

$$\boxed{\Omega_{\Lambda,0} = .692 \pm .012\,.} \tag{3.13}$$

Worked problem 3.3: A ‘force’ that counteracts gravity

We have seen above that a positive cosmological constant leads to an exponentially expanding universe with $a(t) = a_0 e^{H_0(t-t_0)}$, see equation (2.24). Since our universe's energy density is dominate by dark energy with $w = -1$ this should approximate our universe fairly well.
Assume one object is on an approximately circular motion around another (like for example the earth around the sun). If the two objects are a distance $r = a(t)\,d$ apart, the cosmological expansion leads to an outward acceleration $\ddot{r} = \ddot{a}\,d = H_0^2 r$. Compare this acceleration for our value of H_0 to the inward acceleration due to gravity. Plug in values for the systems earth-sun, sun-milky way and milky way-Virgo supercluster (vsc).

Solution: The acceleration on a circular orbit is $v^2/r = v^2/(a(t)d)$. The velocity squared for a light object orbiting a massive object of mass M is given by $v^2 = GM/r$. This leads to the following

	outward: $H_0^2 r$	inward: GM/r^2
earth-sun	$7 \times 10^{-25}\, m/s^2$	$6 \times 10^{-3}\, m/s^2$
sun-milky way	$1 \times 10^{-15}\, m/s^2$	$3 \times 10^{-9}\, m/s^2$
milky way-vsc	$3 \times 10^{-12}\, m/s^2$	$5 \times 10^{-13}\, m/s^2$

We see that dark energy is not strong enough to have a substantial effect on the solar system or galaxies but it can affect superclusters.

So, we have seen that our current universe has negligible curvature and negligible contributions from radiation. The total energy density splits into roughly 70% dark energy (behaving like a cosmological constant) and 30% non-relativistic matter. The fact that most of this non-relativistic matter is made out of unknown particles might come as a big surprise. It often leads to the statement that we don't understand 95% of our universe since we don't understand the dark energy either, however, we know the equations of state parameters for dark matter and dark energy very well, so that we can describe the history of our universe very accurately. As we mentioned above, compared to radiation both dark matter and dark energy become less and less important at earlier times due to their different scaling with $a(t)$, so that they actually are unimportant in our description of the very early universe (the same is true for curvature, i.e., the extra contribution in the first Friedmann eqn. for non-zero K).

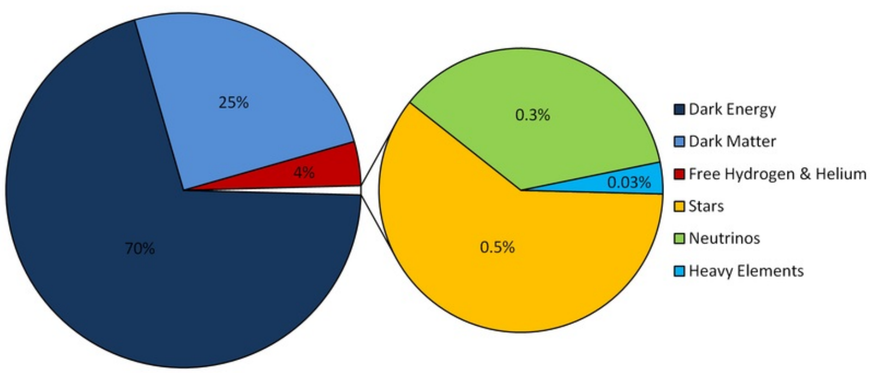


Figure 15: Pie-Chart of the matter and energy content of our universe (taken from Wikipedia).

The different contributions to our universe are summarized in figure 15, whose values slightly differ from the ones I have given above. The reason is that contributions from dark energy and matter can still change at the level of a few percent since they are sensitive to the value of the Hubble constant that has not yet been measured that accurately (recall that this is the reason why we defined it in terms of an unknown h as $H_0 = 100h\, km/(sMpc)$). In the last few years experiments have measured central values of h ranging from .74 to .67. If you think that this is not overly precise you have to keep in mind that in particular satellite experiments take a very long time from the planning stage until the data is analyzed and usually different experiments determine cosmological parameters in very different ways. So, we should be happy that they all seem so close to each other that they are mutually consistent. Also, as we will see in the next section, the discovery of dark energy, that is the *dominating* form of energy in the current universe, was only roughly 20 years ago!

Worked problem 3.4: Starlight

Do a back of the envelope estimate to determine the contribution of starlight to the energy density of our current universe.

Solution: Stars are fusing hydrogen and helium to heavier elements and release energy in the form of starlight. Nuclear fusion converts a few per mil of the rest mass energy of a star into other forms of energy like neutrinos and starlight. Usually there are several fusion processes, which leads to the following estimate

$$\begin{array}{c}\text{rest mass} \cdot \text{few per mil} \cdot \text{several fusion processes} \\ \cdot \text{ fraction that goes into photons not neutrinos}\end{array} \approx 1\% \cdot \text{rest mass}\,. \tag{3.14}$$

As we will learn below, the heavy elements above in figure 15 were created in stars. If we add the entire contribution from all stars that are currently still fusing material to the rest mass, we find an upper limit for starlight of

$$\Omega_{\text{starlight}} \approx (.5\% + .03\%) \cdot 1\% \approx 5.3 \times 10^{-5}\,. \tag{3.15}$$

Clearly this is negligible today. Note that, contrary to the primordial radiation that dominated our early universe, the starlight did not exist in the very early universe (before the first stars and galaxy formed). Hence, the energy density of starlight never played an important role in the evolution of our universe.
Note, that the neutrino contribution in figure 15 is from primordial neutrinos in the comic neutrino background (see below).

3.3 Solving Friedmann's equation for our universe

With the above information we have everything we need to solve Friedmann's equation and determine the scale factor of our universe. In order to do that let us first use the definition of the critical density today $\rho_{c,0} = 3H_0^2/(8\pi G)$ (cf. equation (3.2)) to rewrite the Friedmann equations for $K = 0$ (cf. equations (1.27) and (1.28))

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

$$
\begin{aligned}
&= H_0^2 \left(\Omega_{m,0} \left(\frac{a_0}{a(t)} \right)^3 + \Omega_{r,0} \left(\frac{a_0}{a(t)} \right)^4 + \Omega_{\Lambda,0} \right) , \qquad (3.16) \\
\frac{\ddot{a}(t)}{a(t)} &= -\frac{4\pi G}{3} (\rho(t) + 3p(t)) \\
&= -\frac{H_0^2}{2} \left(\frac{\rho(t) + 3p(t)}{\rho_{c,0}} \right) \\
&= -\frac{H_0^2}{2} \left(\Omega_{m,0} \left(\frac{a_0}{a(t)} \right)^3 + 2\Omega_{r,0} \left(\frac{a_0}{a(t)} \right)^4 - 2\Omega_{\Lambda,0} \right) . \qquad (3.17)
\end{aligned}
$$

It is a straightforward exercise to show that the second equation is not independent from the first, so we can solve only the first Friedmann equation. Unfortunately, there is no simple closed form solution but it can be solved easily numerically with the boundary condition $a(0) = 0$, see figure 16.

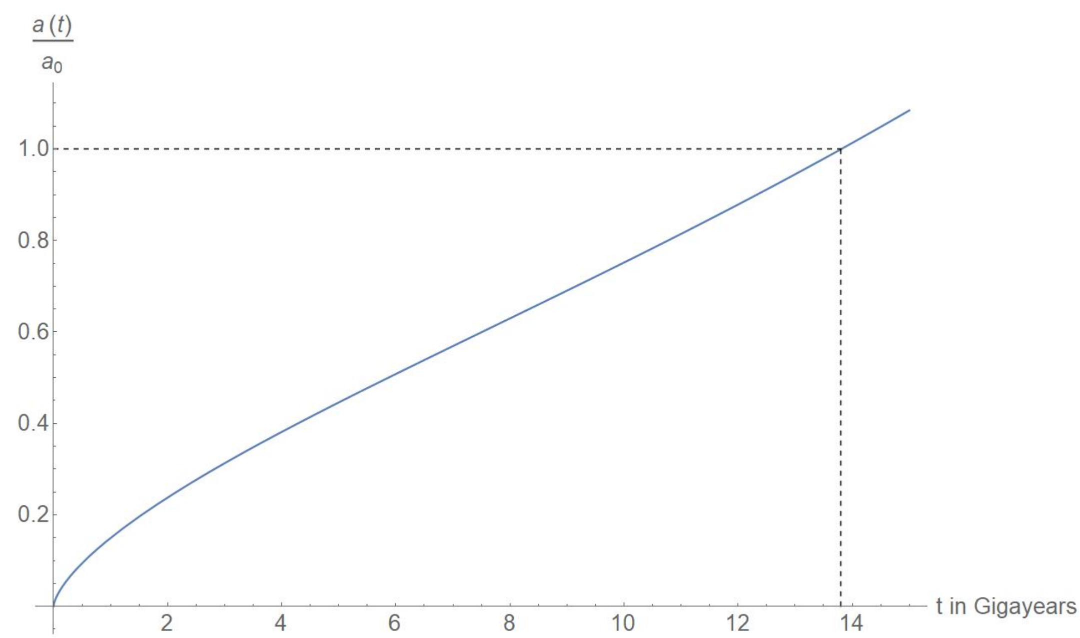


Figure 16: A plot of the scale factor of our universe. The dashed black lines denote today.

From the above numerical solution, we can determine t_0 such that $a(t)/a_0 = 1$, i.e., we can find the age of the universe to be $t_0 = 13.8Gyr$. We can also study different time ranges. For example, for the early universe when $a(t)$ is very small, we expect from the above equation (3.16) that radiation will dominate the evolution. This should then lead to $a(t) \propto \sqrt{t}$, cf. equation (2.26). This is indeed what we can see from our numerical solution in figure 17.

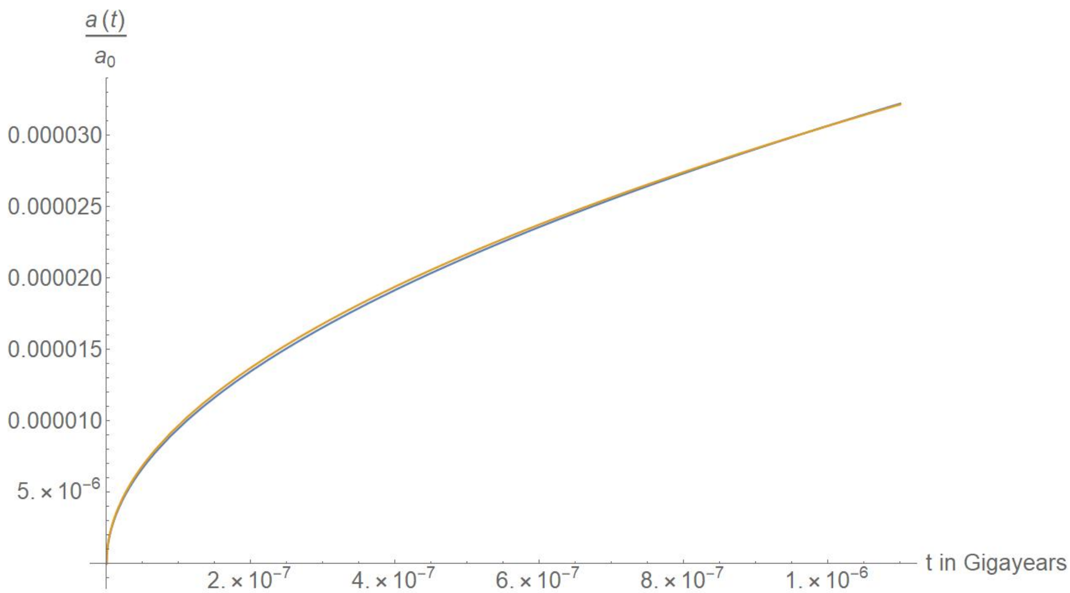


Figure 17: A plot of the scale factor of our early universe (in blue). The orange line that nearly coincides with the blue is given by $\sqrt{t/Gyr}/32.6$.

In an intermediate time range we find that the scale factor behaves like $a(t) \propto t^{\frac{2}{3}}$, which is the expected behavior for a universe dominated by matter, cf. equation (2.27). This is shown in figure18.

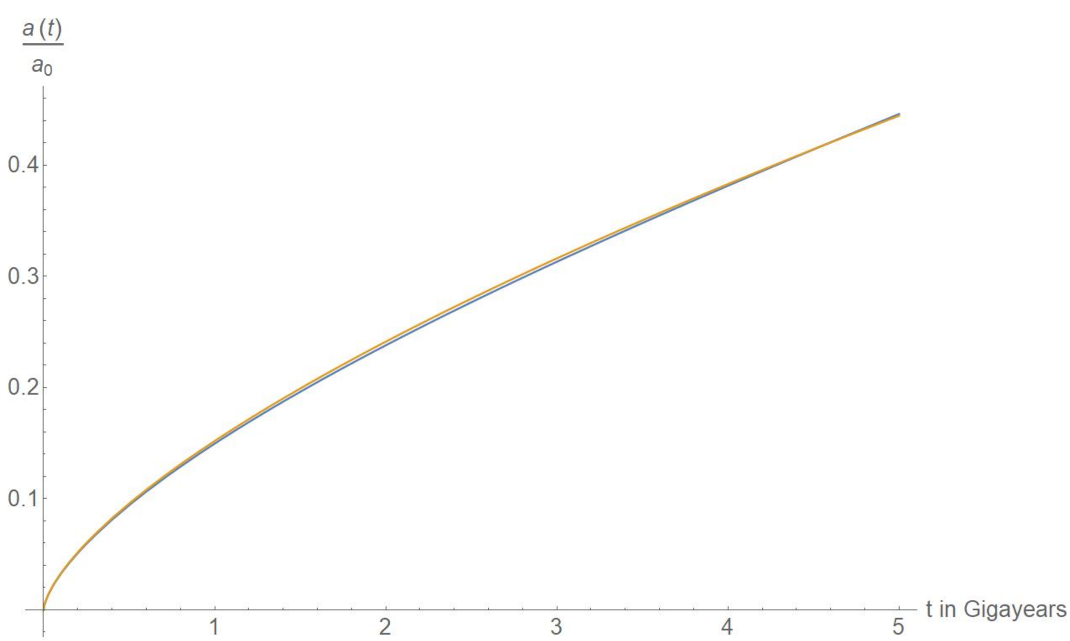


Figure 18: A plot of the scale factor of our universe (in blue). The orange line that nearly coincides with the blue is given by $(t/Gyr)^{\frac{2}{3}}/6.58$.

In the future we expect the universe to expand and therefore the matter should get diluted like $a(t)^3$ while the dark energy should remain constant. This means the expansion should become exponentially in t, as we have derived in equation (2.29). This is not visible in figure 16 although the dark energy is the dominant form of energy in our current universe. However, if we plot the scale factor further into the future, then we can see the exponential expansion, as shown in figure 19.

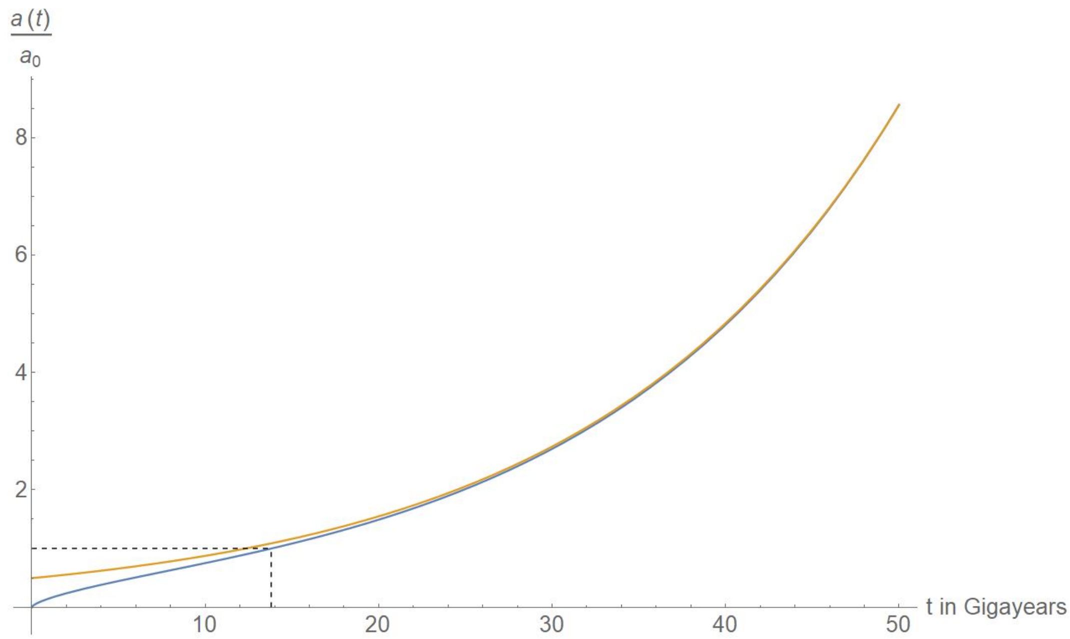


Figure 19: A plot of the scale factor of our universe (in blue). The orange line that coincides with the blue for large t is given by $.495e^{.057t/Gyr}$.

Thus, we see from solving the scale factor for our universe that its evolution can be divided into three different parts: An early epoch that is dominated by radiation, an intermediate epoch dominated by matter and finally the current epoch that is dominated by dark energy. We will return to this in the next section after first studying in more detail dark energy and the fate of our universe in the next subsection.

3.4 Dark energy

Arguably the most important discovery in cosmology in the last twenty-five years is the discovery of dark energy. In 1998 the High-Z Supernova Search Team and in 1999 the Supernova Cosmology Project published their analysis of type Ia supernovae, which are a type of standard candles in cosmology. Their observations are in strong tension with a matter dominated universe and much more compatible with a universe whose expansion is accelerating. For this discovery S. Perlmutter, B. Schmidt and A. Riess were awarded the 2011 Nobel Prize in Physics. Before we look at their data, we need to review several useful definitions.

3.4.1 Redshift

As we have discussed previously, in an expanding universe the wavelength λ of a photon gets stretched as well. For example, if a photon is emitted at time t_1 with wavelength λ_1 and we observe it today at time t_0 with wavelength λ_0, then we have the simple relation

$$\lambda_0 = \frac{a(t_0)}{a(t_1)}\lambda_1 \,. \tag{3.18}$$

Note, in particular for an expanding universe we have $a(t_0) > a(t_1)$ so that the wavelength becomes larger $\lambda_0 > \lambda_1$. This means that the observed light is 'redshifted'. This terminology arises from the fact that for visible light the red wavelengths are the longest. Since bluish light has the shortest wavelengths one likewise uses the term 'blueshift', if wavelength become shorter. This would happen, if the universe contracts or a star is moving towards us with a speed that overcompensates the redshift from the expansion of the universe.

The fractional shift in the wavelength of photons is the so-called redshift parameter

$$\boxed{z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{a(t_0)}{a(t_1)} - 1 \,.} \tag{3.19}$$

Since we know the time evolution of $a(t)$ in our universe, we have a one-to-one map from t to $a(t)$ so we can assign to events either a time t or a particular value of the scale factor $a(t)$ (modulo the overall rescaling for $K = 0$ that we discussed in problem 3.2). Likewise, we can use equation (3.19) to assign a redshift z to a particular time t_1 in the past. For example, currently the oldest observed galaxy has a redshift parameter of $z_G \approx 11$ which corresponds to a time of a little bit more than 420 million years after the big bang. So, for our stars and galaxies the values of z are rather modest, however, for the cosmic microwave background that will play an important role in the following sections the redshift is $z_{CMB} \approx 1000$.

Worked problem 3.5: Behavior of the redshift

(a) Discuss the behavior of the redshift for events in the very early universe when t_1 approaches the big bang.

(b) What is the redshift factor today?

(c) Since our universe is expanding it will be, at some time in the future, twice as big. What will be the redshift at that future time?

(d) What will be the redshift factor in the infinite future in a universe that keeps expanding forever?

Solution:

(a) While t_1 becomes smaller, $a(t_1)$ likewise becomes smaller, approaching zero, and therefore the redshift parameter will diverge, going to positive infinity.

(b) For $t_1 = t_0$ one trivially finds that $z = 0$.

(c) If the universe in the future is twice as big, we have $a(t_1) = 2a(t_0)$. This leads to a redshift

$$z = \frac{a(t_0)}{a(t_1)} - 1 = \frac{1}{2} - 1 = -\frac{1}{2}\,. \tag{3.20}$$

So, while the definition in terms of a photon wavelength does not make sense for future events, we can use the definition of the redshift in terms of the scale factor for past, current and future events. All future events will have negative redshift and all past events will have positive redshift.

(d) In an forever expanding universe $a(t_1) \to \infty$. This means the redshift factor will become $z \to -1$. So, while going back in time the redshift factor can become arbitrarily large, if we extend it into the future of our universe it cannot become arbitrarily large and negative: $-1 < z < \infty$.

As you might have noticed from the definition above in equation (3.19), the redshift parameter tells us how much smaller the universe was when the light was emitted. We find

$$\frac{a(t_1)}{a(t_0)} = \frac{1}{1+z} \quad \Rightarrow \quad \frac{a(t_1)}{a(t_0)} \approx \frac{1}{z} \quad \text{for} \quad z \gg 1\,. \tag{3.21}$$

So, for example, when the first galaxies were formed around $z_G \approx 10$ the universe was roughly one tenth of its current size. When the cosmic microwave background was emitted at $z_{CMB} \approx 1000$ the universe was 1/1000 of its current size.

3.4.2 Accelerated Expansion

The Supernova Cosmology Project studied 42 type Ia supernova with redshift parameter between $z \approx .2$ and $z \approx .9$. The result is shown in figure 20. The plot resembles Hubble's original plot of distance vs. velocity (cf. figures 3 and 4). However, this plot is somewhat different. On the logarithmic x-axis we have the redshift that, as we have seen above, encodes the time in the past when the light was emitted or likewise the distance the light has traveled from the type Ia supernova to us. On the y-axis we see something called "effective m_B" which is the observed brightness. It is related to the observed flux here on earth via $F \propto 10^{-2m_B/5}$.[5] Since the light curves for type Ia supernova are so homogeneous, we know how the observed flux correlates with the distance. The axis are chosen such that a universe with only matter, namely $(\Omega_{m,0}, \Omega_{\Lambda,0}) = (1, 0)$ would lead to a straight line. This straight line is shown in the figure 20 in dashed light blue: the somewhat hidden third blue line from the top behind the middle solid black line (that denotes the same). This is the dust filled universe that we studied above in subsection 2.3. We see that the data points for large redshift on the right of

[5] See for example §1.3-1.6 in Weinberg's "*Cosmology*" book for more details.

the plot are consistently above this line and cluster somewhere between the top two blue lines that denote universes with $(\Omega_{m,0}, \Omega_{\Lambda,0}) = (0.5, 0.5)$ and $(\Omega_{m,0}, \Omega_{\Lambda,0}) = (0, 1)$. So, this is consistent and was the first step towards the more precise values we quoted above for our universe $(\Omega_{m,0}, \Omega_{\Lambda,0}) = (.308, .692)$. Intuitively we can understand the data as follows: In a universe with cosmological constant that pushes things apart the supernovae will be further away than in a universe with only matter. Thus, less of the supernovae luminosities is observed on earth due to their larger distances. This means the photon flux $F \propto 10^{-2m_B/5}$ is smaller and this manifests itself in a larger effective m_B. This effect accumulates with distance so for supernovae that are very close, it is very small and becomes larger and larger for supernovae that are further away.

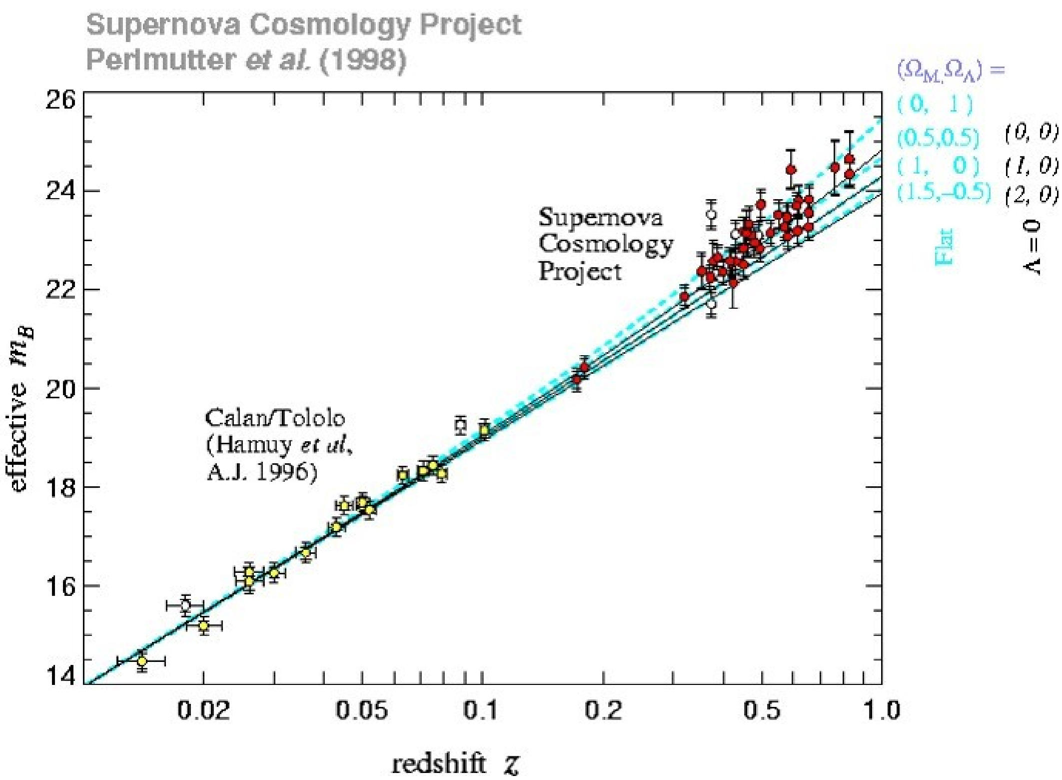


Figure 20: Hubble diagram obtained from 42 high-redshift type Ia supernovae from the Supernova Cosmology Project and 18 low redshift supernovae from the Calan-Tololo Supernova Survey. We see that the observed expansion favors a universe with matter and a cosmological constant.

So, we see that the data favors a universe with a substantial contribution from a cosmological constant. When these supernovae were originally studied it was clear that there was severe tension with a flat matter dominated universe and that any kind of energy density that leads to an accelerated expansion would help to explain the discrepancy. While a cosmological constant is the most natural energy density that does the job, it is not the only possibility. From the second Friedmann equation and using $p(t) = w\rho(t)$ we find that any fluid with $w < -1/3$ leads to an accelerated expansion:

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi G}{3}(\rho(t) + 3p(t)) = -\frac{4\pi G}{3}\rho(t)(1+3w) > 0 \qquad \Leftrightarrow \qquad w < -\frac{1}{3}\,. \quad (3.22)$$

So, the equation of state parameter for the dark energy initially didn't have to be $w = -1$ but could have been very different. However, during the last decade experiments have measured the equation of state parameter for the dark energy very accurately and the current value is[6]

$$w_{DE} = -1.006 \pm 0.045\,. \quad (3.23)$$

This is very much consistent with a cosmological constant.

Assuming that the dark energy has equation of state parameter $w = -1$ one can fit the theoretical predictions of a universe with matter and cosmological constant with the data from supernovae (SNe), the cosmic microwave background (CMB) and baryon acoustic oscillations (BAO) to determine the density parameters as shown in figure 21.

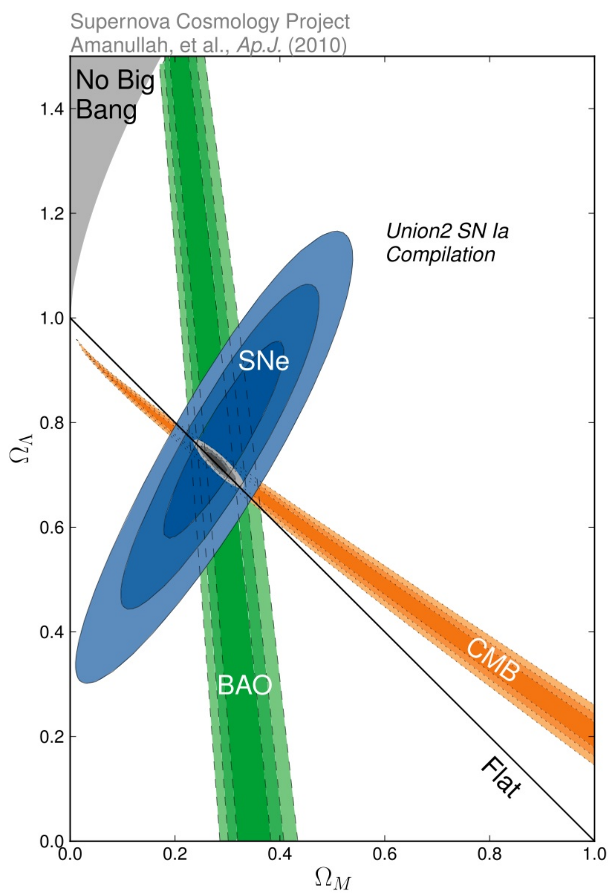


Figure 21: Different experiments exclude different parameter regions leaving only a very small part in the $(\Omega_m, \Omega_\Lambda)$-plane. These three observations are very different and it is very satisfying that all three slices intersect so nicely.

[6]There are slightly different values for this equation of state parameter and in particular the error bar can be substantially larger, depending on which other parameters we hold fixed or let vary.

3.4.3 The smallness of the cosmological constant

Interestingly the value of the cosmological constant differs from its natural value by a factor of 10^{-120}, which is certainly the biggest discrepancy between theoretical expectation and measured value that we have ever observed in nature. As you checked in the first homework, if the cosmological constant would be anywhere near its natural value no structures could have formed and such a universe would be empty and lifeless. It actually turns out that the value of the cosmological constant can't be that much bigger than what we observe, since otherwise structure formation would not take place. Concretely, the cosmological constant can only be larger by a factor of roughly 200 since otherwise hydrogen and helium clouds would not have gravitationally clumped to form stars and galaxies.

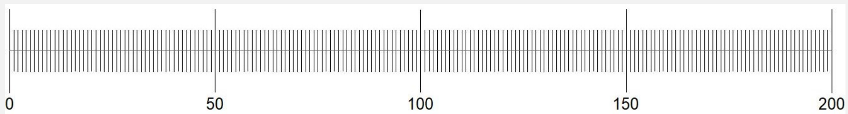
This factor of 200 does not sound too bad compared to 10^{-120}, however, this fact by itself does not explain the smallness of the observed value. For such an explanation we would first have to assume that there are a gigantic number of universes with random values for the cosmological constant and if this were the case then we would of course live in one that allows for lifeforms to exist. If the cosmological constant is somewhat randomly distributed between zero and the Planck scale then we would need of the order of 10^{120} different universes for this explanation to work! While this sounds crazy, it seems actually very likely that our best candidate for a theory of quantum gravity, which is called string theory, does indeed have so many (or actually many more) solutions.

This *anthropic argument* is pretty unsatisfying. Essentially, we didn't explain or derive the smallness of the cosmological constant at all from any underlying theory. However, it might be that this is how things are. A similar problem is the distance between the earth and the sun. If the earth would be much closer or further away, then there would be no liquid water and life as we know it wouldn't exist. Kepler tried to derive the distance between planets and the sun from an underlying theory. Now we know that Newton's theory of gravity or Einstein's general relativity do not constrain the orbits of the planets but these are rather randomly distributed. Since there are a lot of planets in our solar system and we have also discovered plenty of planets orbiting other stars, the *anthropic argument* in this case is rather common sense. A big difference here is that we can observe other planets. If we would not be able to observe any evidence for the existence of other universes, then this theoretical idea couldn't really be verified.

Worked problem 3.6: The factor of 200

In 1987 Steve Weinberg calculated the upper bound on the cosmological constant Λ above which structure formation in the early universe would not be possible due to the exponential expansion of the universe. As we mentioned above this upper bound is roughly $\Lambda_{up} \approx 200\Lambda_{obs}$, where Λ_{obs} is the observed value. This seems spectacular given that the naturally expected value is larger by 10^{120}. However, Weinberg was not happy with this: Assume a large set of different universes with evenly distributed values for the cosmological constant. What is the most likely value for all universe with $\Lambda < \Lambda_{up}$?

Solution:



From the above we see that most of the universes should have a value for the cosmological constant that is of the same order of magnitude as the upper bound Λ_{up}. Only a very small fraction will have much smaller values. So, it seems as if one would still require some fine-tuning to obtain a universe like our own. However, as was pointed out by Vilenkin, at the upper bound structure formation is barely possible (and proceeds very slowly). So, universes like our own might be somewhat more unlikely in the above sample but they lead to much more structure formation and therefore presumably to much more stars and galaxies, increasing the chance for the existence of observers like ourselves.

3.4.4 The fate of our universe

Now that we know that the current evolution of our universe is governed by a very small cosmological constant let us ask what this means for the future of our universe. As we derived in the previous section, we have $\Omega_m \propto a(t)^{-3}$ and $\Omega_\Lambda = const$. Since $a(t)$ is growing, the matter contribution will become more and more unimportant and our universe is currently entering a phase of exponential expansion. As you can see from equation (2.1) in section 2,

$$\rho(t) \propto a(t)^{-3(1+w)}\,, \tag{3.24}$$

any contribution to the energy density with $w > -1$, will become less and less important in an expanding universe. We don't have any reliable theoretical models that lead to an equation of state parameter $w < -1$, so it seems very plausible that our universe keeps exponentially expanding in the future. What does that mean?

The first cosmological implication is that the universe lives infinitely long and alternatives like a big crunch are excluded. Since the cosmological constant is so tiny, its implications are otherwise rather minuscule. Concretely, within one year the distance between two objects increases due to the exponential expansion roughly by a modest 0.00000001%. This is so small that the initial gas clouds of hydrogen and helium could clump and form stars, galaxies and galaxy clusters. As we studied in worked problem 3.3, the 'smaller' structures like our solar system or our galaxy are gravitational bound and will not really experience a different evolution due to the accelerated expansion of our universe. Also, the galaxy cluster that contains the milky way will stay gravitationally bound. However, other galaxy clusters that are far away from ours will be redshifted more and more and will eventually become unobservable. This is a somewhat counterintuitive fact that we will make precise in the next section. Naively one would have expected that we can see more and more of our universe the longer we wait, since light has more time to reach us. But this intuition is wrong in an exponentially expanding universe and, in the future, we will actually see less and less of our universe!

Having discussed the cosmological fate of our universe based on the current observation, you might be curious about more details and phenomena on smaller scales. For example, stars will eventually burn up all hydrogen and helium and our universe will become a dark place. This doesn't conclude the evolution of our universe and if you are interested you can for example consult Wikipedia for the details of the "Future of an expanding universe".

There is a caveat to the above-described fate of our universe. As we will see later, in order to describe inflation, we will introduce a scalar field that moves in a potential. The cosmological constant can be explained by such a scalar field that sits at a minimum of the potential, where the value of the potential at the minimum is the value of the cosmological constant. In this alternative description, we can then ask whether this minimum is a local or a global minimum. If it is a local minimum, then the scalar field could tunnel quantum mechanically to another minimum with a smaller cosmological constant. This cosmological constant could be zero or even negative. Such transitions are highly suppressed but if they would happen in the future, then this would of course change the evolution of the universe.

Summary: Our universe

In this section we learned about the amounts of different energy densities in our universe: Curvature remains so far unmeasurable small and has therefore no impact on the evolution of our universe. This means that the energy density of our universe today is $\rho \approx 10^{-26} kg/m^3$. 70% of that energy density are in the form of dark energy that behaves within experimental errors like a cosmological constant. The remaining 30% are in the form of non-relativistic matter. These 30% split into 25% that are in the form of an unknown dark matter and 5% that are in the form of standard matter: hydrogen and helium clouds, stars, heavier elements and neutrinos. This means more than 80% of the matter in our universe is in the form of unknown dark matter!
Next, we solved for the scale factor $a(t)$ for our universe and found that its evolution can describe by three eras: First a radiation dominated part, then a matter dominated part and finally the current and future era that is dominated by the dark energy that leads to an exponential expansion.
We defined the redshift $z = \frac{a(t_0)}{a(t_1)} - 1$ and studied in depth the accelerated expansion of our universe that is driven by dark energy. We discussed the cosmological constant problem, i.e., the question of why the amount of dark energy is as small as it is. Due to this dark energy our future universe is expected to expand forever and all matter gets further and further diluted. This means our universe is expected to eventually become a dark and empty space (unless the sign of the dark energy is changing in the future).

4 Our universe from 3.8×10^5 to 13.8×10^9 years

In this section we introduce different commonly used distances in cosmology and we calculate in particular the size of our visible universe. Then we discuss the cosmic microwave background and the overall evolution of our universe from 380,000 years after the big bang until today.

4.1 Particle and event horizon

Let us return to the FRW metric

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

We argued that isotropy forbids mixed terms between the time and spatial coordinates. Then by redefining the time coordinate we can choose the coefficient of dt^2 to be unity. However, there is another very convenient time coordinate, that is called conformal time and we will denote it by τ. It is defined such that

$$dt^2 = a(\tau)^2 d\tau^2 . \tag{4.2}$$

This means that the FRW metric takes the form

$$ds^2 = a(\tau)^2 \left(-d\tau^2 + \frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2 \right) . \tag{4.3}$$

Note, that we chose $K \in \{-1, 0, 1\}$, so that it is clear from (4.3) that r, θ, ϕ are dimensionless, while t and $a(t)$ have the dimension of length or time (recall that $c = 1$). This then means that τ is also dimensionless. Likewise, $\dot{a}(t)$ is dimensionless and

$$a'(\tau) \equiv \frac{da}{d\tau} , \tag{4.4}$$

has the dimension of length or time.

For example, for $K = 0$ the metric (4.3) is just the flat space Minkowski metric multiplied by an overall factor $a(\tau)$. Such a factor that multiplies the entire metric is called a conformal factor, hence the name conformal time for τ. As we have seen in the first lecture the coordinates r, θ, ϕ do not give us *physical distances* since they neglect the factor $a(t)^2$ in the metric. The coordinates r, θ, ϕ are called *comoving coordinates*. Many observable objects like 'standard candles' have a non-zero velocity in comoving coordinates: $\vec{v}_{comoving} = a(t)\dot{\vec{r}}$ in addition to the velocity due to the Hubble expansion $\vec{v}_{Hubble} = \dot{a}(t)\vec{r}$. For far away objects the Hubble velocity is usually much larger while for close by objects like for example cepheids in our galaxy, the Hubble velocity is negligible.

Worked problem 4.1: Comoving vs. Hubble velocities

Stars in galaxies and galaxies within clusters have comoving velocities of a few hundred km/s. Determine the distance at which the Hubble velocity is $v_{Hubble} = 100 km/s$.

Solution: The simplest way to do that is to use Hubble's law as given in equation (1.19)

$$v_{Hubble} = H_0\, d = 67.7 km/s/Mpc\; d = 100 km/s\,. \tag{4.5}$$

This gives a distance of $d = 1.5$ Mpc. In Hubble's original plot in figure 3, he studied galaxies and galaxy cluster up to a distance of 2 Mpc. So, the comoving velocity is responsible for the substantial scattering. However, while the Hubble velocity points away from us, the comoving velocity can be towards us, away from us or perpendicular to our line of sight. Therefore, one can see the Hubble expansion in his original plot.

Light plays a special role in observations but also in determining the causal structure of our universe since no information can travel faster than light. So, two places that cannot exchange light in the lifetime of our universe are causally disconnected. Light follows a null-geodesic which means that $ds = 0$. From the way we have written the metric in equation (4.3), we see that in this case the scale factor $a(\tau)$ does not matter at all and for example for a radially traveling light ray we have

$$ds = 0 \qquad \Rightarrow \qquad d\tau = \frac{dr}{\sqrt{1 - Kr^2}}\,, \tag{4.6}$$

independent of $a(\tau)$.

4.1.1 The particle horizon

Similarly, to a black hole, where the event horizon indicates the horizon beyond which observers from the outside cannot see, i.e., from beyond which they cannot receive any light, there are two important horizons in cosmology. The first horizon, which is called *particle horizon* defines the maximal distance a photon can have traveled since the beginning of the universe. In an expanding universe we have to be precise by what we mean by this distance: We mean the current distance at time t_0 or τ_0 between the photon and the object that emitted it at the beginning of the universe. This is shown in figure 22.

Without loss of generality, we can look at a photon starting at the origin $r = 0$ and traveling outward. So, we have

$$\boxed{d_H(t) \equiv a(\tau) \int_0^{r_H} \frac{dr}{\sqrt{1 - Kr^2}} = a(\tau) \int_{\tau_i}^{\tau} d\tau' = a(t) \int_{t_i}^{t} \frac{dt'}{a(t')}\,,} \tag{4.7}$$

where we used equation (4.6) and then equation (4.3).

For example, for a matter or radiation dominated universe we have $a(t) = a_0 (t/t_0)^p$ with $p < 1$ and the beginning of the universe is at $t_i = 0$. This leads to

$$d_H(t_0) = a_0 \int_0^{t_0} \frac{dt'\, t_0^p (t')^{-p}}{a_0} = \frac{t_0}{1 - p} < \infty\,. \tag{4.8}$$

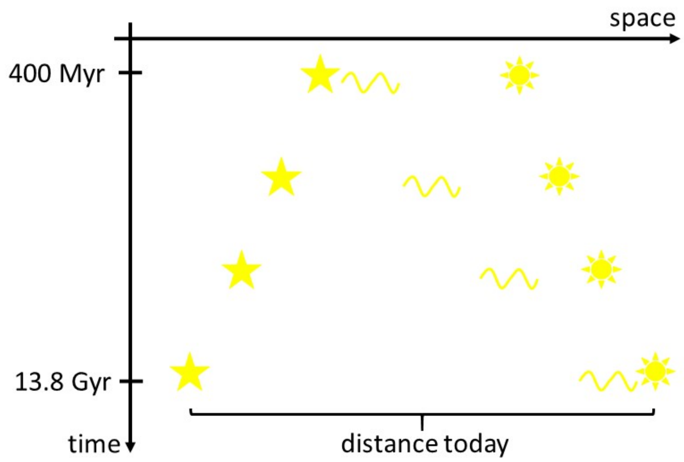


Figure 22: Light is being emitted in the early universe at $t = 400$ Myr by the star on the left. While the light travels to our sun on the right, the distance between the two stars increases due to the expansion of our universe. The particle horizon is the distance to an imaginary star that emitted light at the beginning of the universe, i.e., it is the *size of the visible universe.*

This means that light can have only traveled a finite distance since the beginning of the universe, which is what we would have naively expected. This of course also means that we can only see a finite part of our universe.

For a universe that is exponentially expanding due to a cosmological constant we have $a(t) = a_0 e^{H(t-t_0)}$ and the beginning of the universe, i.e., $a(t_i) = 0$, is at $t_i = -\infty$. This leads to

$$d_H(t_0) = \int_{-\infty}^{t_0} dt' e^{-H(t'-t_0)} = -\frac{1}{H} e^{-H(t'-t_0)} \Big|_{t'=-\infty}^{t_0} = +\infty \,, \tag{4.9}$$

so that in this case the particle horizon is infinite. This fact will be tremendously important once we discuss inflation. The reason is that the cosmic microwave background, which was created shortly after the big bang, is essentially the same on distances much larger than the particle horizon of a matter or radiation dominated universe. This seems in contradiction with causality and requires us to postulate a phase of exponential expansion at the beginning of the universe, which is called inflation.

This cosmic microwave background is the first light in our universe that we can still observe today. It originated shortly after the big bang so the light has been traveling for 13.8 $Gyrs$. We can now ask how big the visible universe is today by calculating the particle horizon

$$d_H(t_0) = a_0 \int_0^{t_0} \frac{dt'}{a(t')} \,. \tag{4.10}$$

For that purpose, it is sufficient to take the matter and the cosmological constant into account which leads to

$$a(t) = a_0 \left(\frac{\Omega_{m,0}}{\Omega_{\Lambda,0}}\right)^{\frac{1}{3}} \left[\sinh\left(\frac{3}{2} H_0 \sqrt{\Omega_{\Lambda,0}} t\right)\right]^{\frac{2}{3}} \,. \tag{4.11}$$

Plugging in the values from the last lecture $\Omega_{\Lambda,0} = .692$ and $\Omega_{m,0} = .308$ leads to

$$d_H(t_0) \approx 3.4t_0 \approx 46.5Gly\,. \tag{4.12}$$

So, our visible universe has currently a radius of $46.5Gly$, although it is only 13.8 $Gyrs$ old. This shouldn't be too surprising since we know that our universe has been constantly expanding since the big bang (cf. figure 22).

Worked problem 4.2: Particle horizon in a dust filed universe

Assume that our universe had no dark energy, i.e., take $\Omega_{\Lambda,0} = 0$ and $\Omega_{m,0} = 1$. How much smaller would the particle horizon be, if the age of the universe is still $t_0 = 13.8$ Gyr?

Solution: For a matter dominated universe, we calculated the particle horizon in equation (4.8), where we need to use that $p = 2/3$,

$$d_H(t_0) = \frac{t_0}{1-p} = 3t_0 = 41.4Gly\,. \tag{4.13}$$

So, without dark energy that pushes things apart faster the event horizon would be smaller by more than $5Gly$.

4.1.2 The event horizon

Another important horizon in cosmology is called the *event horizon*. It refers to the maximal distance light emitted today at t_0 can travel. This horizon determines which parts of space we can exchange information with. If the event horizon is finite, then there are parts of the universe which are causally disconnected from us and similar to the black hole, these parts cannot send information to us (and contrary to the black hole, we cannot send information to these parts of the universe either).

The definition of the event horizon is

$$\boxed{d_e \equiv a_0 \int_{t_0}^{\infty} \frac{dt'}{a(t')}\,.} \tag{4.14}$$

Let us again first look at a matter or radiation dominated universe with $a(t) = a_0(t/t_0)^p$ and $p < 1$. We find

$$d_e = \int_{t_0}^{\infty} \frac{dt'\, t_0^p}{t'^p} = \left.\frac{t_0^p\, t'^{(1-p)}}{1-p}\right|_{t'=t_0}^{\infty} = \infty\,. \tag{4.15}$$

Again, this result is consistent with our naive expectation. In the infinite time until the end of the universe the light can travel an infinite distance, so that in such a universe we could send and receive signals from anywhere in the universe. However, as we discussed in the previous section, our universe is currently and, in the future, dominated by a cosmological constant and $a(t)$ approaches an exponential expansion $a(t) = a_0 e^{H(t-t_0)}$. This means that we will *never* be able to see the entire universe

$$d_e = \int_{t_0}^{\infty} dt'\, e^{-H(t'-t_0)} = -\frac{1}{H} e^{-H(t'-t_0)}\Big|_{t'=t_0}^{\infty} = \frac{1}{H} = const\,. \tag{4.16}$$

So, we see that even if the light travels infinitely long it can only tell us about places at a finite distance. Intuitively we can understand this since the exponential expansion constantly stretches the space between two objects. If the distant is larger than $1/H$, then in any given amount of time the increase of the distant due to the stretching is larger than the distant light can travel.

Worked problem 4.3: Evolution of Hubble parameter

Use the numerical solution for $a(t)$ for our universe that was discussed in subsection 3.3 to find the Hubble value in the far distant future.

Solution: Using the numerical solution for $a(t)$ it is trivial to find $H = \dot{a}/a$, which is plotted below.

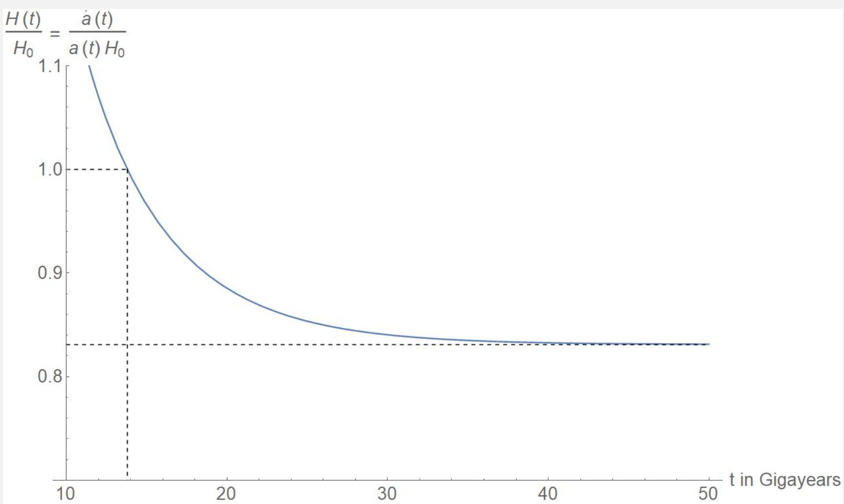


Figure 23: A plot of the time evolution of the Hubble parameter in our universe.

From the above we see that the Hubble parameter is shrinking with time. It asymptotes to the value $H(\infty) = .83 H_0 \approx 1/17.4 Gyr \approx 56.3 \frac{km}{s\,Mpc}$.

Since our universe has currently still a substantial amount of matter, the Hubble parameter is changing until in the far future the energy density is almost completely given by the dark energy and the Hubble parameter becomes constant. This asymptotic Hubble parameter in our universe is actually not that different from our current value

$H(t=\infty) \approx .83 H_0$, see problem 4.3. So, this means that in the far future we can only exchange information with objects that have a distance of less than $1/H(\infty) \approx 17.4 Gly$. How does that compare to the current event horizon? A lower bound would be $1/H_0 \approx 14.4 Gly$. However, matter that is still important in our current universe leads to a larger event horizon. Using the solution $a(t)$ given in equation (4.11) and that describe our current and future universe very well, we find the event horizon for our universe to be

$$d_e = a_0 \int_{t_0}^{\infty} \frac{dt'}{a(t')} \approx 16.6 Gly\,. \tag{4.17}$$

Since during an exponential expansion objects that are not gravitationally bound to our galaxy will move further and further away from us, they will actually leave our event horizon in the future. This means for example that very distant galaxy clusters that we can exchange information with today will at some point in the future leave our event horizon and become unreachable. As mentioned before, contrary to intuition, we will therefore be able to exchange information with less parts of our universe in the future.

We can also ask how much of the universe we can ever observe in the future. We have seen that the first light (the CMB) has a current distance that is far bigger than the above event horizon. This is due to the fact that our universe in the past wasn't dominated by dark energy and therefore didn't have a finite even horizon. Once we have a finite event horizon, only light within this horizon can reach us. This means that there is light from far distant objects that has entered our event horizon already but hasn't reached us yet. The maximal distance *today* of such objects that we will be able to see in the future can be calculated as sum of the particle and event horizon

$$d_H(t_0) + d_e = a_0 \int_{0}^{\infty} \frac{dt'}{a(t')} \approx 4.6 t_0 \approx 64 Glyrs\,, \tag{4.18}$$

where we used $a(t)$ from equation (4.11).

The current event horizon and particle horizon for our universe are shown in figure 24 as beige and blue regions. The light from the light blue region with radius $d_H(t_0) + d_e(t_0) \approx 64$ Glyr will still reach us in the future because it has already entered our event horizon. So, the *current* size of the visible universe that we will see in the far distant future is 64 Glyr.

In the far future the universe keeps expanding exponentially. That leads to an event horizon $d_e(t_f) = 1/H(t_f) = 17.4$ Gyr (cf. worked problem 4.3). The visible universe with radius $d_H(t_f)$ will grow exponentially since the prefactor in front of the integral is $a(t_f) \propto e^{H(t_f)t_f}$ and the integral itself is finite and will approximate $d_H(t_0)+d_e(t_0) \approx 64$ Glyr, see figure 25. However, the part of the universe that we cannot see that is outside the visible universe will also grow exponential and we will actually not see much more in the far distant future. While the event horizon will take on a fixed finite value in the future, we will be able to explore less and less of the universe in the future. For example, a reachable galaxy that is currently 10 *Gly* away would most likely at some point in the future leave our even horizon due to the exponential expansion of the universe and thereby become unreachable.

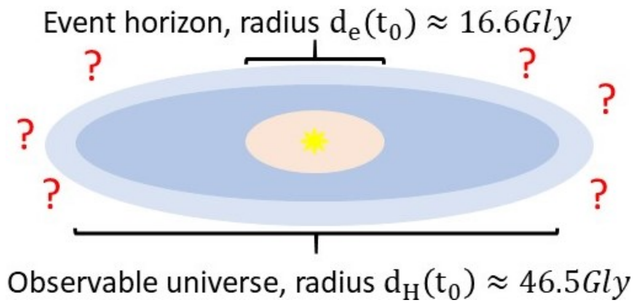


Figure 24: Summary of the particle and event horizon in our current universe with $t_0 = 13.8$ Gyr. The sun denotes us and the (beige) event horizon denotes the region that we can reach with light signals emitted today. The observable universe in darker blue is what we can actually see today out of the potentially much larger universe. Light from the light blue region with radius $d_e(t_0) + d_H(t_0)$ has already entered our event horizon and we will be able to see it in the future. The red question marks indicate regions that we know nothing about.

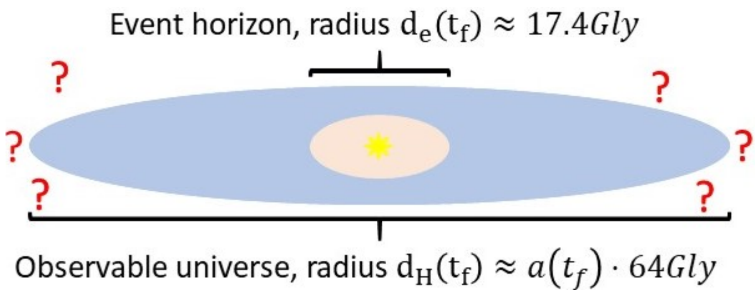


Figure 25: The event horizon will be a little bit larger in the far future at $t_f \gg 13.8$ Gyr. The light blue region will continuously shrink while the darker blue visible universe will grow eventually exponentially like $a(t_f) \propto e^{H(t_f)t_f}$. However, the unknown region with red question marks will likewise grow exponentially and we will never see more than the light blue region above in figure 24 that is now the darker blue region in the far future.

4.2 The cosmic microwave background (CMB)

In 1964 Arno Penzias and Robert Wilson were working on the detection of radio waves that bounced off echo balloon satellites when they discovered a faint background of radiation in the microwave range using the antenna shown in figure 26.

Surprisingly this signal seemed to come from everywhere in the sky. Checking their antenna, they discovered a family of nesting pigeons that they removed together with what Penzias called "white dielectric material" (aka bird poop). Nevertheless, the signal remained. At the same time some astrophysicists were planning to search for such a signal since they had realized that, if the universe had started in a hot dense state, then the subsequent expansion would lead to photons whose wavelength would get red-shifted due to the expansion of the universe in such a way that their wavelength is today in the millimeter or micrometer range. This is exactly what Penzias and Wilson

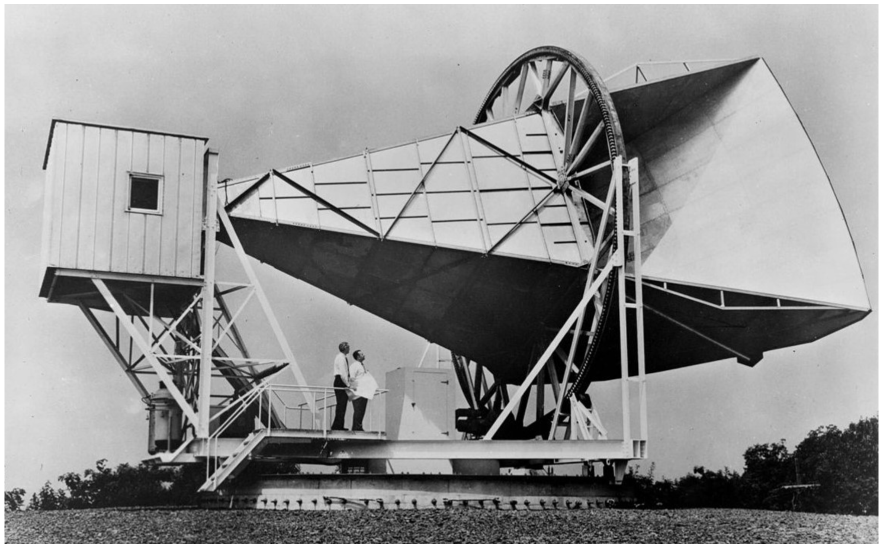


Figure 26: The Holmdel Horn Antenna in use in 1962 by Penzias and Wilson.

discovered and for this they were awarded the 1978 Nobel Prize.

In the subsequent decades these photons that are called the cosmic microwave background (CMB) since they fill the universe homogeneously and isotropically have been studied by many ground-based and satellite experiments. This CMB is a clear evidence for a hot big bang and at the same time the best tool for precision measurements in cosmology. The COBE satellite that was launched in 1989 was the first space-based experiment that measured the CMB. It showed that the CMB follows the best black-body spectrum ever observed in nature (see figure 27).

George Smoot and John Mather, two of COBE's principal investigators, were awarded the 2006 Nobel Prize in physics for their work on the COBE project. This shows the great importance of the CMB in understanding the evolution of our universe from the very beginning until today.

Recall that the radiation from a black body is described by Planck's law which gives for $c = \hbar = 1$

$$B(\lambda, T) = \frac{4\pi}{\lambda^5} \frac{1}{e^{\frac{2\pi}{\lambda k_B T}} - 1} \,. \tag{4.19}$$

Here $k_B = 8.6 \times 10^{-5} eV/K$ is the Boltzmann constant given in terms of electron volts and Kelvin. The black body spectrum of the CMB corresponds to a temperature of

$$T_{CMB,0} = 2.72548 \pm 0.00057K \approx -270°C \,. \tag{4.20}$$

So, we see that the CMB is pretty cold.

Worked problem 4.4: Wavelength of CMB photons

What is the wavelength of photons from the cosmic microwave background at its peak intensity?

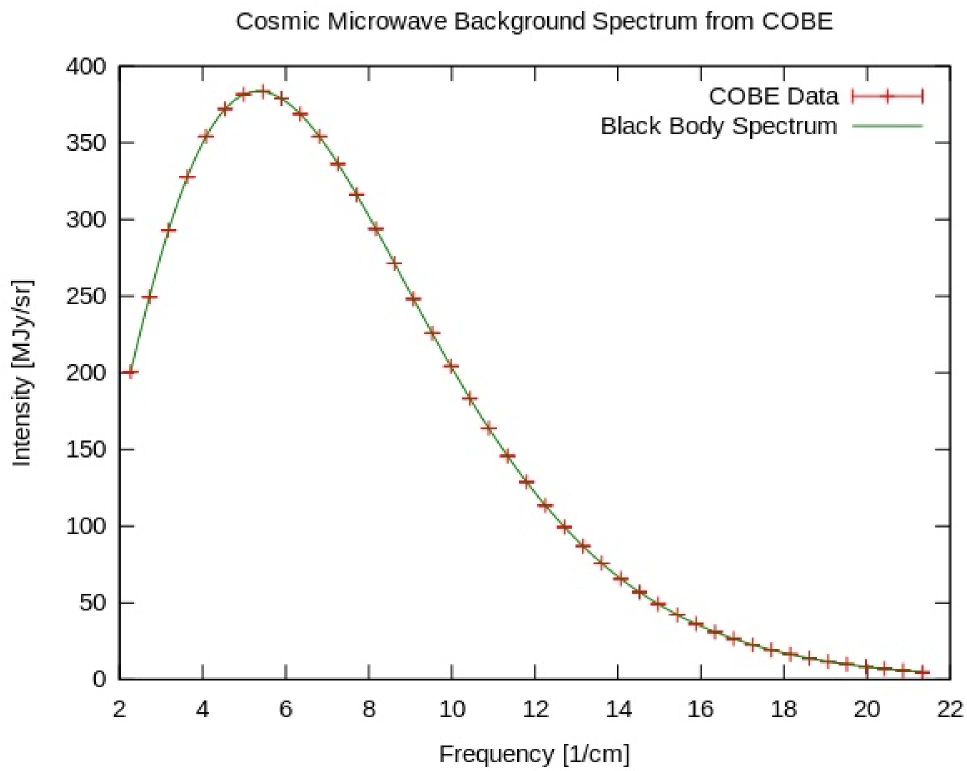


Figure 27: The black body spectrum of the CMB as measured by the COBE satellite. The errors bars are too small to observe and it is impossible to distinguish the theoretical curve from the measured spectrum (figure taken from Wikipedia).

Solution: From Wien's displacement law we find

$$\lambda_{max} = 2898\mu m\, K/T \approx 1mm\,. \tag{4.21}$$

The maybe confusing thing to note is that microwaves do not have wavelengths of micrometers but rather wavelengths between 1 millimeter and 1 meter.

In an expanding universe the wavelength of photons gets red-shifted so that a current wavelength λ_0 was at an earlier time t_1 $\lambda_1 = a(t_1)\lambda_0/a_0$. This tells us that the wavelength scales like $a(t)$ and then it follows from (4.19) that the temperature of the CMB scales like $1/a(t)$. In particular the temperature of the CMB at the earlier time t_1 is given by

$$T_{CMB}(t_1) = \frac{a_0}{a(t_1)} T_{CMB,0} = (1+z)T_{CMB,0}\,. \tag{4.22}$$

So, we see that the temperature of the CMB was much larger for a much smaller universe. This means that the further we go back in time the hotter the universe was.

Worked problem 4.5: Temperature of the CMB

We mentioned above that the CMB photons were emitted at a redshift of $z_{CMB} \approx 1000$. What was the temperature at that time and what is $k_B T$ in eV?

Solution: The temperature is trivially given by

$$T_{CMB}(t_{CMB}) = (1+1000)T_{CMB,0} = 2727K\,. \tag{4.23}$$

Now using that $k_B = 8.616 \times 10^{-5} eV/K$ we find an energy $E = k_B T = .2eV$. As we will discuss in much more detail below, at this point the black body spectrum contains already a substantial number of photons with energies around 13.6 eV. So, we are at the transition between ionized and atomic hydrogen.

We have seen in the last lecture that the CMB radiation is only contributing a very small amount to the current energy density of our universe. Nevertheless, the radiation energy density has the strongest dependence on the scale factor and will therefore inevitably dominate in the very early universe. Recall from the last lecture that the Friedmann equation for our universe can be written as

$$\rho_c(t) = \rho_\Lambda + \rho_m(t) + \rho_{rad}(t) \approx \rho_c(t_0)\left[.7 + .3\left(\frac{a_0}{a(t)}\right)^3 + 10^{-4}\left(\frac{a_0}{a(t)}\right)^4\right]. \tag{4.24}$$

While the cosmological constant is currently and, in the future, dominating the energy density, this was different at earlier times when $a(t)$ was much smaller than a_0. In particular, if we plot the energy density as a function of a as a log-log-plot we find the following history of our universe:

While we are currently (and in the future) in an era dominated by the dark energy, this was different in the past. The substantial amount of matter in the universe was dominating its evolution until fairly recently. In the far distant past, when the universe was much, much smaller, radiation was actually the dominating form of energy density since it grows like $a(t)^{-4}$.

Recall that a curvature contribution proportional to $K/a(t)^2$ was never the dominating form of energy density. It is currently very small and will become less important in the future since it decays with increasing $a(t)$ while the dark energy is most likely constant. Going backwards in time the matter and radiation contributions will grow faster than the curvature so that the curvature was less important in the past. So, the curvature was never dominating but it could nevertheless be non-zero and measurable with more precise experiments in the future. This would then not really affect the part of the evolution of our universe that we are currently discussing but it would be interesting on theoretical grounds and might hint at or exclude certain transitions in the very, very early universe.

Above we have argued that the temperature of the CMB is decreasing over time and therefore in the past the universe was much hotter. From equation (4.24) and the

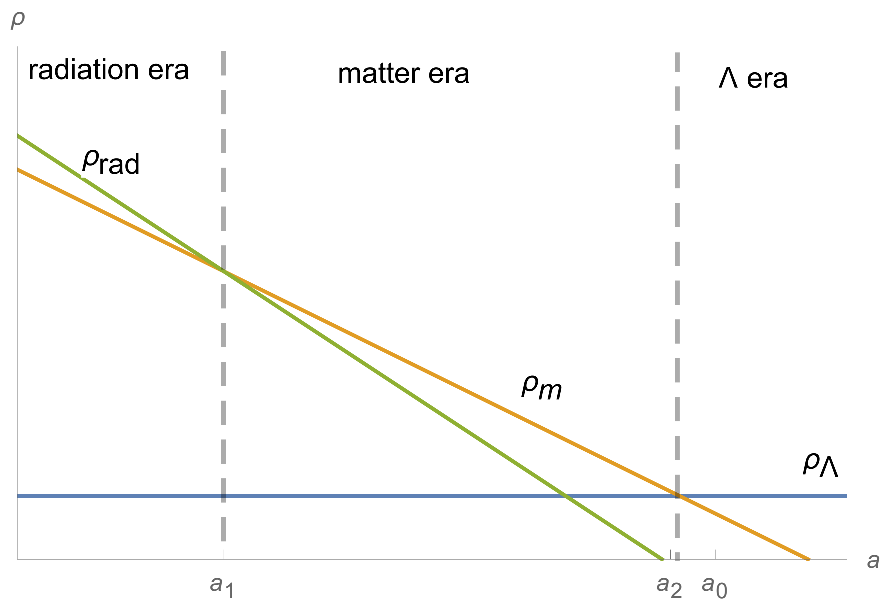


Figure 28: The evolution of our universe.

figure 28 we see that the universe in the past also had a much higher energy density. So, our universe started out in a much hotter and much denser state that then got diluted and cooled due to the expansion of the universe. This means that we can use thermodynamics and our knowledge of particle and nuclear physics to understand its early evolution.

4.3 From 380,000 years after the big bang until today

Before we delve into the more involved evolution of the early universe, let us first discuss the evolution from the time the CMB was released until today. The following picture shows the rather few important cosmological periods of our universe since the release of the cosmic microwave background:

As discussed before and studied in the homework, the cosmological constant started to dominate the evolution of the scale factor $a(t)$ a few billion years ago. However, there is substantial amount of baryonic and dark matter in the universe and this form of non-relativistic matter was dominating in the not-too-distant past.

During this matter dominated era very small deviations from a perfectly homogeneous universe were amplified by gravity and structures started to form. We will later discuss these inhomogeneities in more detail and understand their amazing origin: quantum fluctuations! These very small effects in our everyday life have actually led to the small inhomogeneities that are the seeds of our stars and galaxies, which is one of the most amazing features of our universe. As we will discuss in the next sections a few hundred years after the big bang our universe was a soup of atomic nuclei, electrons and photons. The photons constantly interacted with the electrons via Compton scattering and the negatively charged electrons interacted with the positively charged nuclei via the Coulomb force. Atoms like hydrogen were not stable as long as there were photons with an energy larger than the binding energy of hydrogen

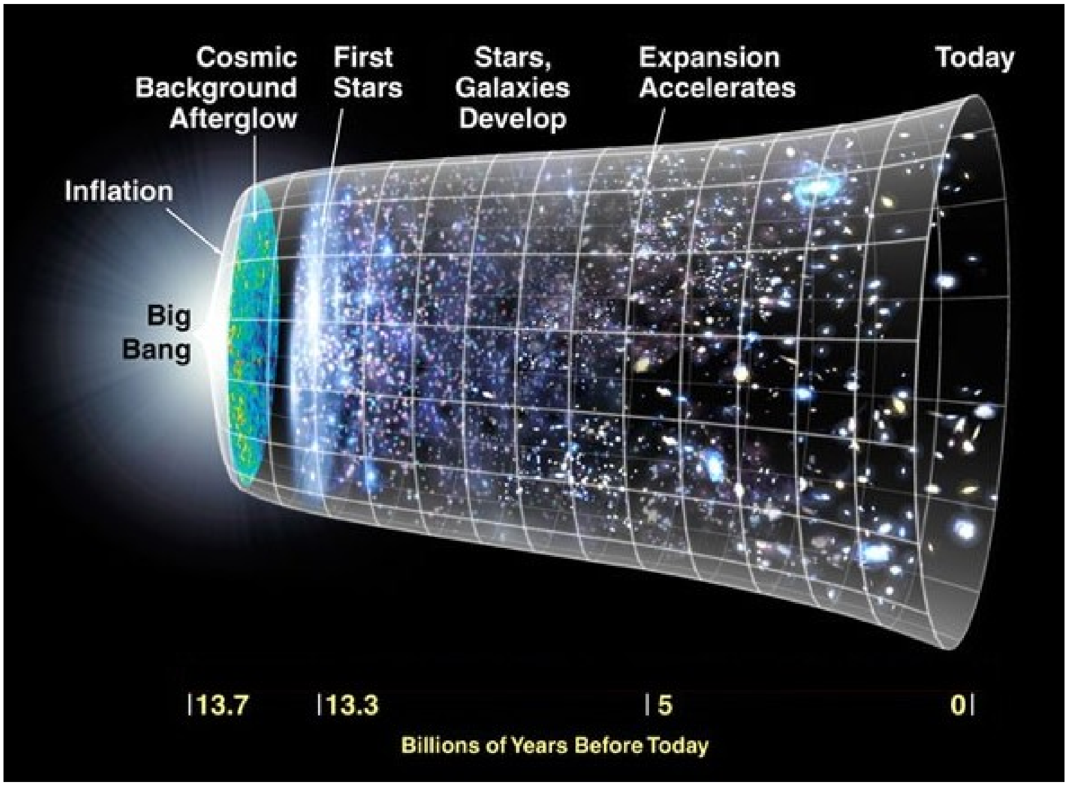


Figure 29: The evolution of our universe since the release of the cosmic microwave background 3.8×10^5 years after the big bang.

which is $13.6eV$, since these photons would ionize the hydrogen. However, at around 380,000 years after the big bang the universe had cooled enough so that stable atoms could form. The photons at this time had mostly energies below the $13.6eV$ threshold and could not ionize the atoms anymore. Since the atoms are electrically neutral their interaction with the photons became negligible at this point and the photons could essentially stream freely. These photons are what we observe today so they have been traveling for more than 1.3×10^{10} years! Detailed studies of the photons and particular the deviation from homogeneity and isotropy tell us a lot about the universe at the time of the so called 'last scattering' after which these photons were able to stream freely. However, these photons also tell us about much earlier times as we will see later in this course.

As mentioned above, when neutral atoms formed roughly $380,000$ years after the big bang, very small inhomogeneities started to amplify due to the gravitational attraction: The initially more or less homogeneous distribution of neutral atoms contained only hydrogen, helium and lithium. Any small inhomogeneity in such a setup will amplify due to gravity: Denser regions will attract more matter and become even denser and therefore increase their gravitational attraction. Such regions that become more and more dense will eventually after a few hundred million years lead to the first stars. In these stars the gravitational attraction is sufficiently strong to start nuclear fusion, which provides the energy for the stars to shine and also leads to the creation of elements heavier than Lithium. These heavier elements were then released at the end of the lifetime of the first stars in supernovae explosions. Stars and planets like the ones in our solar system that are then later on formed contain heavier elements. This star formation process out of the initial hydrogen and helium mixed with some heavier elements will go on for roughly 10^{14} years. This follows since as we have discussed in the previous section, a large amount of hydrogen and helium is still in clouds outside of stars.

Summary: Distances and overview of our universe and its fate

We introduce the formal definition of the particle horizon d_H and the event horizon d_e

$$
\begin{aligned}
d_H(t_0) &= a_0 \int_0^{t_0} \frac{dt'}{a(t')}\,, \\
d_e(t_0) &= a_0 \int_{t_0}^{\infty} \frac{dt'}{a(t')}\,.
\end{aligned}
\tag{4.25}
$$

The particle horizon denotes the radius of the visible universe toady and is given by 47 Gly. This is substantially larger than the naive estimate based on the universe's age of 13.8 Gyr because the universe is expanding. The surprising fact about our universe is that, even in the infinite far future, the particle horizon will remain finite: 67 Gly. The reason for that is the exponential expansion due to the dark energy.
The event horizon denotes the maximal distance a light signal released today can travel. Again, surprisingly this is finite in our universe and only 16.6 Gly. This is again due to the exponential expansion caused by the dark energy. The event horizon will slightly increase in the future to an asymptotic value of 17.4 Gly since the Hubble parameter will decrease. While no information, light signal or alien can reach us from a distance larger than the event horizon, very distant galaxies and galaxy cluster will move outside of our event horizon due to the Hubble expansion of our universe. This means that in the far distant future we can interact with less galaxies instead of more. However, our galaxy and local cluster that are gravitationally bound cannot be moved apart by the dark energy.
Next, we discussed the cosmic microwave background (CMB), the earliest visible light in the universe that file our universe. The spectrum follows almost perfectly the blackbody curve with a temperature of $T_{CMB,0} = 2.725K$. Going backwards in time this means that our universe was much denser and hotter since $a(t)$ was much smaller. We have seen that the universe can be roughly divided into three ears: an early radiation dominated era, followed by a matter dominated era and finally today we are in an era of dark energy domination. We also learned that the CMB was released 380,000 years after the big bang when the universe transitioned from a hot plasma to becoming electrically neutral. Then it took a few hundred million years for the primordial hydrogen and helium to form the first stars and galaxies. We expect this star formation to continue for 10^{14} years since there is still a substantial fraction in gas clouds.

Part II - The thermal universe

5 The particles in our universe

In the last section we have shown that our very early universe was in a very hot and dense state. During the expansion of the universe this hot 'soup' cooled and underwent a variety of interesting transitions. In the next few sections, we will discuss the thermodynamical evolution of our universe from a split second after the big bang until the release of the cosmic microwave background 380,000 years after the big bang.

We will from now on set the Boltzmann constant equal to one, $k_B = 1$, and measure temperatures in eV.

5.1 The Standard Model of Particle Physics

Currently the world's largest particle accelerator, the Large Hadron Collider, does experiments at an energy of up to 13 TeV. So, we understand particle physics up to this energy scale very well. The particles with masses below this scale and their interactions are described by the so-called standard model of particle physics, which is a particular quantum field theory.[7] Here we will recall some features of the standard model of particle physics and summarize the relevant terminology before we discuss particle physics in the early universe. Compared to the periodic table that you are probably familiar with from your chemistry class, the content of the standard model of particle physics is fairly simple. We know in total of 13 different particles that constitute all the matter in the standard model of particle physics. These 13 particles interact via three different forces: The electro-magnetic force, the weak force and the strong force, which we will all discuss below.

5.1.1 The (known) particles in our universe

The 13 particles come in three different groups:

1. The six *leptons* are probably the particles you are most familiar with. They are fermions and have spin $\frac{1}{2}$. They consist of the electron and its cousins the μ- and τ-particle that all carry one unit of negative electric charge. Additionally, there are three neutrinos that are called ν_e, ν_μ and ν_τ and that all are electrically neutral. All of these six leptons carry no charge under the strong force but they do interact via the weak force. (We will discuss these forces shortly in subsection 5.1.2 below.)
2. There are six more fermionic spin $\frac{1}{2}$ particles that are called *quarks*. These quarks combine to form the probably more familiar protons and neutrons as well as other particles that we will discuss below. The quarks also come in two groups of three particles: the up, charm and top quarks carry $+\frac{2}{3}$ units of electrical charge and the down, strange and bottom quarks carry $-\frac{1}{3}$ unit of electric charge. All quarks are charged under the strong force and the weak force.

[7] Quantum field theories are theories that combine quantum mechanics with special relativity.

3. Lastly there is one more particle in the standard model that was predicted a long time ago but only recently discovered in 2012, the Higgs boson which has spin-0. The Higgs particle plays a special role among all particles. It is responsible for giving a mass to the other particles. This process of giving a mass involves a phase transition in our very early universe. This phase transition changes the cosmological constant we discussed above by a term that is of the order $\lambda_{Higgs} \approx -10^{-65}$. This means that the value of the cosmological constant before this transition λ_{before} needs to cancel with λ_{Higgs} precisely to 55 digits so that $\lambda_{today} \approx 10^{-120} \approx \lambda_{before} + \lambda_{Higgs}$.

All these particles and their masses are shown in figure 30.

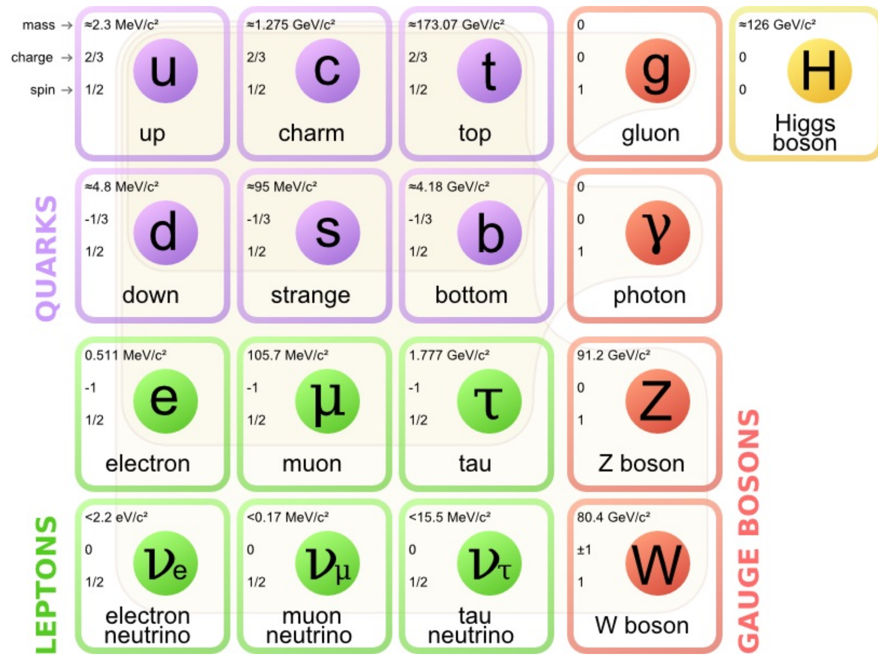


Figure 30: The known particles in our universe (taken from Wikipedia).

Interestingly the leptons and quarks come both in three families (the first three columns in figure 30). Each family contains four particles with the same charges but with different masses. Since the heavier particles in the second and third family can decay into the lighter particles in the first family, it turns out that essentially all standard model particles in our universe are the particles from the first column.[8] This means in particular that all the elements in the periodic table are made up of only of three different particles: the electron, and the up and down quarks.

[8]Neutrinos can oscillate between different families and heavier particles can be created in processes that involve energies larger than their rest mass but these heavier particles quickly decay to up and down quarks and/or electrons.

Worked problem 5.1: Particles with integer charges

You might be confused because quarks carry a non-integer charge, while more familiar particles like the electron, proton and neutron all have integer charges. Is it possible to combine quarks to obtain the electric charges 0 and +1? How many quarks does one at least need?

Solution: We will discuss this below in detail but from the figure 30 above we see that quarks have electric charge -1/3 or +2/3. So, if we want to get 0, we need at least three quarks: One quark with charge +2/3 and two with -1/3. Likewise, if we want to get +1, we need at least three quarks: Two quarks with charge +2/3 and one with -1/3.
As we discuss below, there are also anti-quarks with electric charges -2/3 and +1/3 which leads to further possibilities.

5.1.2 The three forces in the standard model of particle physics

There are four particles that control the interactions between these 13 particles. The three different interactions in the standard model of particle physics are mediated by bosonic spin 1 particles and we quickly discuss the three interactions:

The electro-magnetic force

The force you are probably most familiar with is the electro-magnetic force that is mediated by photons. Many of the particles in the standard model carry an electric charge and therefore interact with photons. Since the photon is massless the interaction strength between two particles falls off as a power law (versus an exponential). This means that the electro-magnetic force could compete with gravity which has the same power law fall off and is in some sense much weaker. However, our universe is essentially electrically neutral on all but very small scales. So, for example, the earth and the sun carry essentially no net electric charge so their interaction is only determined by their masses, i.e., by gravity, and the electro-magnetic force plays no role. The same is true on larger scales that are relevant for cosmology.

The weak force

The weak interaction is probably the force you are most unfamiliar with. It is mediated by three different particles: the electrically neutral Z boson and the two electrically charged $W^{\pm}$ bosons, where the later carry ± 1 unit of electric charge. Since these three particles that mediate the weak force are rather heavy, the weak force is relevant only on fairly short distances, smaller than $10^{-16}m$, like inside a nucleus. There the weak force is responsible for radioactive decay of nuclei.

Worked problem 5.2: A force mediated by massive particles

The weak force is mediated by massive particles with a mass of roughly $m \approx 10^{-25}kg$. Having set $c = \hbar = 1$ a mass can be rewritten as an inverse length $L = 1/m$. What is this length in meters?

Solution: We just need to insert the appropriate powers of $\hbar$ and c to convert the inverse mass into meters

$$L = \frac{1}{m} \approx \frac{1}{10^{-25}kg}\frac{\hbar}{c} \approx \frac{10^{25}}{kg}\frac{10^{-34}m^2kg/s}{3\times 10^8 m/s} = 6.7\times 10^{-18}m\,. \qquad (5.1)$$

If we have set $h = 2\pi\hbar$ to one then we would have found $2\pi L = 4\times 10^{-17}m$. This is the value that actually sets the range of the weak force. This force goes exponentially fast to zero for larger distances.

The strong force

The strong force is what keeps for example the positively charged protons together inside the nuclei. So, it is clear that it is much stronger than the electric force between these charged particles. At low energies the strong force becomes so strong that charged particles cannot exist in isolation. Whenever we try to separate two particles that are charged under the strong force, the energy in the field between the particles becomes so large that it can lead to pair production of new particles that combine with the particles we tried to separate into particles that are neutral under the strong force. The strong force is different from the electromagnetic force in the sense that there are three different types of charges (plus their anti-charges carried by the anti-particles). So, we cannot label the charges by a positive or negative number but rather we have to use three different labels (and anti-labels). For lack of a better label people call the three charges red, green and blue (and the corresponding anti-charges anti-red, anti-green and anti-blue). All quarks carry a single-color charge, i.e., they come in three types so that for example we have a red, a green and a blue up quark. To form particles that are neutral under the strong force we need all three colors to appear once or we can combine a color with an anti-color.

5.1.3 Hadrons, baryons and mesons

Since the quarks are not playing much of a role in our everyday life, let us discuss a little bit more how they form more familiar composite particles, which also allows us to introduce a little bit more terminology. Above we discussed the six leptons, the electron, muon and tau and the three corresponding neutrinos. The name lepton is derived from a Greek word that means fine, thin, little, which is appropriate since, as far as we know, these particles are fundamental in the sense that they are not composed of other particles. The leptons are supposed to be contrasted with the *hadrons*, derived from the Greek word for thick and strong. These are not fundamental particles but rather particles that are kept together by the strong force. These hadrons are furthermore divided into *mesons*, which are bosonic particles with integer spin and *baryons*, which are fermionic particles with half integer spin.

Let us try to build some baryons. As we have heard above, the quarks have to appear in color neutral bound states at low energies. So, we cannot get a baryon that is made up of a single quark. In order to have half integer spin, i.e., a baryon, we therefore need three quarks (and no anti-quarks). If we take three quarks of the same type, like for example three up quarks, then they cannot combine due to the fermi statistics. Since the quarks are spin $\frac{1}{2}$ particles we can combine at most one spin $+\frac{1}{2}$ quark and one spin $-\frac{1}{2}$ quark of the same type. So, the lightest baryon is a composite particle of two up quarks and one down quark, often denoted as uud. This baryon has electric charge +1 and is called the proton. One might think that each of the three quarks can have different color charges so there should be more than one proton, however, this is not the case since color neutrality requires us to take the antisymmetric combination of all color neutral combinations of the three quarks. The proton is as far as we know the only stable baryon. The next heavier baryon is udd. It is the electrically neutral neutron, that has a mean lifetime of slightly less than 15 minutes. These two baryons will play a very important role in the creation of nuclei in our very early universe, as we will discuss in section 8. All other baryons have a very short lifetime so that they quickly decay into protons and neutrons.

The mesons are all unstable so they do not play such an important role in cosmology. Let us nevertheless discuss the pions that are the lightest mesons: We want to get a particle with integer spin so we need (at least) two quarks to construct a meson. Since we want these particles to be color neutral, we need actually a quark and an anti-quark to construct the simplest and lightest mesons. Restricting again to hadrons formed out of the u and d quarks and anti-quarks we seem to have four possibilities to construct two quark mesons: $u\bar{u}$, $d\bar{d}$, $u\bar{d}$ and $\bar{u}d$, where a bar over a quark denotes the anti-quark. However, the actual meson particles we observe are sometimes linear combinations of quark-anti-quark pairs. In particular the sum of $u\bar{u}$ and $d\bar{d}$ combines with $\pm s\bar{s}$ to form two η mesons. This leaves us with only three light π mesons that are made up of u and d quarks and anti-quarks: The $\pi^0 = (u\bar{u} - d\bar{d})/\sqrt{2}$, the $\pi^+ = u\bar{d}$ and the $\pi^- = \bar{u}d$, where the superscript denotes the electric charges in units of the electron charge. All these mesons decay very quickly with a mean lifetime of $2.6 \times 10^{-8} s$ for the $\pi^\pm$ and $8.4 \times 10^{-17} s$ for the π^0.

Worked problem 5.3: The masses of hadrons

The masses of the proton and the neutron are $m_p = 938 MeV/c^2$ and $m_n = 940 MeV/c^2$. Compare these to the sums of the masses of their constituent quarks.

Solution: From figure 30 we can read off $m_u = 2.3 MeV/c^2$ and $m_d = 4.8 MeV/c^2$. This leads to

$$\begin{aligned} 2m_u + m_d &= 9.4 MeV/c^2 &\quad \leftrightarrow \quad & m_p = 938 MeV/c^2 \\ m_u + 2m_d &= 11.9 MeV/c^2 &\quad \leftrightarrow \quad & m_n = 940 MeV/c^2 \end{aligned}$$

Clearly, these masses are very different and this difference is due to the binding energy that is actually almost ten times as large as rest mass energy of the constituent particles.

5.1.4 Gravity and a theory of everything

Note, that the standard model or particle physics neglects gravity entirely. This is very well justified in most regimes of interest to elementary particle physics but it tells us that in order to describe our entire universe, we need another theory that unifies quantum field theories with gravity in a so-called theory of everything.

General relativity that we are using in this course to describe the evolution of our universe is likewise incomplete since it is a classical theory and it inevitably breaks down near the Planck scale $1/\sqrt{G} = 1.22 \times 10^{19} GeV$. We will only use the so-called *reduced* Planck mass which is given by [9]

$$M_P = \frac{1}{\sqrt{8\pi G}} = 2.435 \times 10^{18} GeV \,. \tag{5.2}$$

Since the universe was at higher and higher temperatures/energies the further back in time we are going, we reach a point at which general relativity cannot be used anymore. This in particular means we cannot use general relativity to understand the actual beginning of our universe, i.e., the big bang. It is also unclear whether we will ever be able to get experimental insight into the physics that caused the big bang. However, between the energies of less than an eV at the time the CMB was released until the energies studied in particle accelerators of a few TeV we have 12 orders of magnitude of well understood physics to discuss and from the TeV range up to the Planck scale we will discuss another 15 orders of magnitude in energy of slightly more speculative physics. There are also ideas for theories of quantum gravity that go beyond general relativity and might allow us to theoretically understand the initial singularity that arises in general relativity at the beginning of our universe.

Worked problem 5.4: The Planck time

Calculate the Planck time by recalling that an inverse mass or energy can be rewritten as a time.

Solution: We have seen before in worked problem 5.2 that we can convert a mass into a length so with another appropriate insertion (or actual removal) of $c = 1$, which connects length and time, we can also convert it into a time:

$$\frac{1}{M_P} = \frac{\hbar}{2.435 \times 10^{18} GeV} = \frac{6.58 \times 10^{-16} eV \cdot s}{2.435 \times 10^{18} GeV} = 2.7 \times 10^{-43} s \,. \tag{5.3}$$

In principle we would therefore hope that we might be able to use general

[9] We will often drop for simplicity the word reduced and refer imprecisely to M_P as the Planck mass.

relativity to describe our universe at time $t > 2.7 \times 10^{-43}s$ after the big bang. However, it is not fully clear at what scale general relativity breaks down. Strictly speaking we also cannot use it to even discuss the time range between the start of the universe and the time general relativity becomes applicable.

5.2 The universe in thermal equilibrium

In the early universe at temperatures above a few hundred GeV all standard model particles will have energies that are much larger than their rest mass:

$$E(p) = \sqrt{m^2 + p^2} \approx p\,. \tag{5.4}$$

This means that they do not behave like non-relativistic (pressureless) matter but rather like radiation (i.e., like for example photons for which $E = p$). Since the masses are negligible in this era, there is only one scale in the standard model which is the rate of interactions Γ, i.e., the number of interactions per time. In principle this rate can be different for the different particles but we neglect this for the rough estimates in this section. In our expanding universe there is one more length or time scale set by the Hubble scale H. If the particles interact a lot without feeling the expansion of the universe, then they will be in local equilibrium. This would mean that

$$\Gamma \gg H\,. \tag{5.5}$$

If the above equation is true, then we can use equilibrium thermodynamics to describe our universe. We would therefore like to estimate during which temperatures/energies the above is expected to be true.

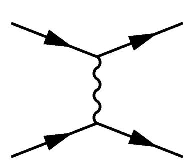
The particle interaction rate can be written as

$$\Gamma \equiv n\sigma v\,, \tag{5.6}$$

where n is the number density, i.e., the number of particles per volume, σ is the interaction cross-section and v is the average velocity of the particles. Since, as we argued above, all particles are highly relativistic for $T \gg 100GeV$, we have $v \approx c = 1$. The only dimensionful quantity is the temperature T, that has the dimension of an energy which is the same as an inverse length. So, we find for the number density and the cross section

$$n \sim T^3\,, \qquad \sigma \sim T^{-2}\,. \tag{5.7}$$

For the cross section we can be more precise. Two particles interact dominantly via the exchange of one of the gauge bosons (that are all massless above $100GeV$). We often write this in terms of Feynman diagrams that use straight lines to indicate fermions and wiggly lines to describe a gauge boson:



The interaction cross section is the square of such a diagram so that it goes like the fourth power of the interaction strength between the fermion and the gauge boson. This interaction strength is usually called $\sqrt{\alpha}$, which gives

$$\sigma \sim \frac{\alpha^2}{T^2}\,. \tag{5.8}$$

Putting this together we find the following scaling of the interaction rate

$$\Gamma \approx n\sigma \approx \alpha^2 T\,. \tag{5.9}$$

The actual value of α depends on the particular force with which the particles interact as well as the energy scale. However, at high energies the interaction strengths of all forces seem to become almost the same, which hints at a unification of all force in a so called grand unified theory (GUT). At this GUT scale the energy is approximately $10^{16}GeV$ and the value of α is $\alpha \approx .05$.

As we have seen above, our early universe was dominated by radiation.[10] This means that

$$H^2 = \frac{\rho}{3M_P^2} \sim \frac{1}{a(t)^4 M_P^2} \sim \frac{T^4}{M_P^2} \qquad \Rightarrow \qquad H \sim \frac{T^2}{M_P}\,. \tag{5.10}$$

Putting this together we find that

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_P}{T} \sim \frac{10^{16}GeV}{T}\,, \tag{5.11}$$

which means for roughly $100GeV \ll T \ll 10^{16}GeV$ we have $\Gamma \gg H$. So, our early and hot universe was in a state of local equilibrium and we can describe it using equilibrium thermodynamics.

5.3 Baryogenesis

In the following we will describe the cooling of our universe starting from a 'soup' of matter and photons at a temperature of a few hundred GeV, using our knowledge of particle physics and thermodynamics. However, before we do that let us mention a puzzle: In a very hot universe we can create particle-anti-particle pairs from photons, denoted γ. For example, for the electron e^- and the positron e^+ we can have the reversible process

$$e^- + e^+ \leftrightarrow \gamma + \gamma\,. \tag{5.12}$$

In an expanding universe we know that the photons 'lose' energy due to the redshift, $E \propto 1/a(t)$. This means that there is a certain point at which the two photons on the right won't have enough energy to create an electron-positron pair. At that moment the above process should only go in one direction

$$e^- + e^+ \rightarrow \gamma + \gamma\,. \tag{5.13}$$

If the early universe has an equal number of particles and anti-particles then eventually, we would expect that all particles and anti-particles annihilate and leave a universe

[10] This is also plausible from the above discussion that showed that all the standard model particles behaved like radiation instead of matter in the early universe.

filled with photons. However, in our universe there is an asymmetry between matter and anti-matter, so that our universe ended up with matter and not just radiation. This asymmetry can be quantified by the ratio between the number of baryons (protons and neutrons) and photons in our current universe. Observations tell us that today

$$\frac{n_b}{n_\gamma} \approx 6 \times 10^{-10}\,, \tag{5.14}$$

while the same ratio for anti-baryons seems to be essentially zero. There is no mechanism inside the standard model of particle physics that can explain this so called baryogenesis, i.e., the observed matter-anti-matter asymmetry, so we will simply assume that the initial conditions of the universe were such that they lead to the observed baryon to photon ratio.[11]

Worked problem 5.5: Baryon asymmetry

Assume a particular volume of the early universe contained n_q quarks and $n_{\bar{q}}$ anti-quarks. During the cooling process each quark-anti-quark pair annihilates and produces two photons. Determine the smallest integers n_q and $n_{\bar{q}}$ that lead to the above ratio in equation (5.14). You can use that approximately $n_q \approx n_{\bar{q}}$.

Solution: We have $n_q \approx n_{\bar{q}}$ and since each quark-anti-quark pair annihilates into two photons we have $n_\gamma = n_q + n_{\bar{q}} \approx 2n_q \approx 2n_{\bar{q}}$. We also need to recall that each baryon contains three quarks. So, after the quark-anti-quark annihilation we are left with $n_b = (n_q - n_{\bar{q}})/3$. This leads to

$$\frac{n_b}{n_\gamma} = \frac{(n_q - n_{\bar{q}})/3}{2n_q} = \frac{n_q - n_{\bar{q}}}{6n_q} \approx 6 \times 10^{-10}\,. \tag{5.15}$$

To get the smallest integers let us assume that a single baryon survives: $n_b = 1 = (n_q - n_{\bar{q}})/3$. This leads to

$$\frac{1}{2n_q} \approx 6 \times 10^{-10}\,. \tag{5.16}$$

The solution is $n_q = 833,333,333$ which implies $n_{\bar{q}} = 833,333,330$. So, we see that in the early universe the different between matter particles and anti-matter particles is absolutely tiny.
We assumed above that the initial state has only quarks and anti-quarks but in reality, there are also photons in the very early universe. However, their number is comparable to the number of quarks and anti-quarks $n_\gamma \sim n_q \sim n_{\bar{q}}$, so that the above calculation wouldn't really change by much.

Note that the asymmetry in the early universe has to exist not only in baryons but also in leptons like the electron since ultimately electrons and the baryons combine to

[11] There are a variety of theoretical ideas of how such an asymmetry can arise but so far the experiments have not singled out any particular model, so we refrain from discussing baryogenesis in any further detail.

form the atoms in our universe. In the early universe there were ample interactions that also converted quark-anti-quark pairs into electron-positron pairs

$$q + \bar{q} \leftrightarrow e^- + e^+ \,. \tag{5.17}$$

This means in turn that it is not absolutely necessary, that baryogenesis that creates an asymmetry between baryons and anti-baryons, had to happen in our universe. It is also possible that there was *leptogenesis* that led to an asymmetry between leptons, like the electron e^-, compared with anti-leptons, like the positron e^+. Due to the ample interactions in the early universe either baryogenesis or leptogenesis or a combination of both could have led to the observed asymmetry today.

Summary: Particle physics in the early universe

Compared to the periodic table in chemistry, particle physics with 17 fundamental particles is rather simple. Almost all the regular matter in our universe (and everything in the chemistry periodic table) is made out of only three particles: the electron and the up and down quarks. The latter two combine to form the more familiar protons and neutrons.
The particles in our universe interact via the electromagnetic, the weak and the strong forces as well as gravity. Our very early universe was a hot, dense soup of fundamental particles that were due to ample interactions in thermal equilibrium. This opens up the possibility to use equilibrium thermodynamics to describe the evolution of our early universe analytically in the next sections.
One of the remaining puzzles in early universe cosmology is the difference in matter and anti-matter: While the initial soup of particles and anti-particles cools in the expanding universe, particles and anti-particles annihilate each other. This ultimately led to our universe that contains no more anti-matter but some regular matter that formed the stars and galaxies. There is however no process in the Standard Model of Particle Physics that can explain this asymmetry between matter and anti-matter in the early universe. So, new physics is needed to give rise to this so-called baryogenesis.

6 The thermal universe

In the previous section we have seen that the standard model particles in the early universe were interacting so much that the Hubble expansion is negligible compared to the interaction rate, while the temperature is in the range $100GeV \ll T \ll 10^{16}GeV$. This means that there is ample time for the standard model particles to be in thermal equilibrium by the time the temperature is a few hundred GeV. We can therefore use equilibrium thermodynamics to discuss the evolution of this soup of standard model particles as the universe expands and cools.

6.1 Equilibrium Thermodynamics

In order to understand the number density n, energy density ρ and pressure P [12] for different particles in the early universe we need to know their distribution as a function of phase space, i.e., their distribution in real space and momentum space encoded in a function $f(\vec{r}, \vec{p})$. For a homogeneous distribution, this phase space function cannot depend on the spatial coordinate $\vec{r}$ and for an isotropic distribution the phase space function can only depend on the absolute value of the momentum $p = |\vec{p}|$. For a system of particles in equilibrium the distribution function is given by the Fermi-Dirac distribution for fermions and by the Bose-Einstein distribution for bosons. Both can be written as

$$f_{\pm}(p) = \frac{1}{e^{(E(p)-\mu)/T} \pm 1}\,, \tag{6.1}$$

where the $+$ sign is for fermions and the $-$ sign is for bosons and μ denotes the chemical potential that generically is a function of the temperature T.

Due to the ample interactions in the early universe all particles have the same average kinetic energy, i.e., the same temperature T so that we do not need to keep track of different temperatures. In the early universe the chemical potentials for all the particles are small so that we can neglect them and set $\mu = 0$. However, this would mean that the number of particles and anti-particles is the same which isn't quite true as discussed above in subsection 5.3 on baryogenesis. A small non-zero chemical potential allows one to account for the small matter-anti-matter asymmetry but it substantially complicates the analysis and is not needed for our discussion.

Allowing for g internal degrees of freedom, i.e., for particles with spin, the particle density in phase space is given by $\frac{g}{(2\pi)^3} f(p)$, where we dropped the subscript $\pm$ to avoid cluttering.[13] In order to obtain the number density n, we need to integrate this over the momentum

$$n \equiv \frac{g}{(2\pi)^3} \int d^3p f(p)\,. \tag{6.2}$$

To obtain the energy density ρ, we need to weigh each state by its energy $E(p) = \sqrt{m^2 + p^2}$ so that we have [14]

$$\rho \equiv \frac{g}{(2\pi)^3} \int d^3p f(p) E(p)\,. \tag{6.3}$$

Lastly the pressure is defined as

$$P \equiv \frac{g}{(2\pi)^3} \int d^3p f(p) \frac{p^2}{3E(p)}\,. \tag{6.4}$$

The integrals in n, ρ and P have to be evaluated numerically unless we are in

[12] We will switch conventions and denote the pressure by P to avoid confusion with the absolute value of the momentum $p = |\vec{p}|$.

[13] For massless particles $g = 1$ for real scalar fields and $g = 2$ otherwise. For massive particles the spin S determines g via $g = 2S + 1$.

[14] For strongly interacting particles we would have to take into account the interaction energy, but the particles in the early universe were weakly interacting so that we can neglect the interaction energy.

particular limits. For such limits we will make use of the general formulas

$$\int_0^\infty du \frac{u^n}{e^u - 1} = \zeta(n+1)\Gamma(n+1)\,, \tag{6.5}$$

$$\int_0^\infty du u^n e^{-u^2} = \frac{1}{2}\Gamma\left(\tfrac{1}{2}(n+1)\right)\,, \tag{6.6}$$

where ζ is the Riemann zeta-function and the Γ-function is an extension of the factorial function and in particular takes the values $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}^*$.

6.1.1 The relativistic limit

Let us first evaluate n, ρ and P for relativistic particles:

$$E(p) = \sqrt{m^2 + p^2} \approx p \gg m\,. \tag{6.7}$$

We define $y = p/T$ so that $f_\pm(y) = 1/(e^y \pm 1)$. For bosons we then find

$$n_b = \frac{g}{(2\pi)^3} \int_0^\infty \frac{4\pi T^3 y^2 dy}{e^y - 1} = \frac{gT^3 \zeta(3)\Gamma(3)}{2\pi^2} = \frac{\zeta(3)}{\pi^2} g\,T^3\,, \tag{6.8}$$

where $\zeta(3) \approx 1.2$. For fermions we can use that

$$\frac{1}{e^y + 1} = \frac{1}{e^y - 1} - \frac{2}{e^{2y} - 1}\,, \tag{6.9}$$

to get

$$n_f = n_b - \frac{g}{(2\pi)^3} \int_0^\infty \frac{8\pi T^3 y^2 dy}{e^{2y} - 1} = n_b - \frac{g}{(2\pi)^3} \int_0^\infty \frac{\pi T^3 \tilde{y}^2 d\tilde{y}}{e^{\tilde{y}} - 1} = n_b - \frac{1}{4} n_b = \frac{3\zeta(3)}{4\pi^2} g\,T^3\,. \tag{6.10}$$

So, we have found for relativistic particles that

$$n_b = \frac{4}{3} n_f = \frac{\zeta(3)}{\pi^2} g\,T^3\,. \tag{6.11}$$

Note, that the scaling of T^3 agrees with our scaling assumption in equation (5.7) in section 5.

Now let us likewise calculate the energy density

$$\rho_b = \frac{g}{(2\pi)^3} \int_0^\infty \frac{4\pi T^4 y^3 dy}{e^y - 1} = \frac{g}{2\pi^2} T^4 \zeta(4)\Gamma(4) = \frac{\pi^2}{30} g T^4\,, \tag{6.12}$$

where we used that $\zeta(4) = \pi^4/90$. For fermions we find

$$\rho_f = \rho_b - \frac{g}{(2\pi)^3} \int_0^\infty \frac{8\pi T^4 y^3 dy}{e^{2y} - 1} = \rho_b - \frac{g}{(2\pi)^3} \int_0^\infty \frac{1}{2} \frac{\pi T^4 \tilde{y}^3 d\tilde{y}}{e^{\tilde{y}} - 1} = \rho_b - \frac{1}{8}\rho_b = \frac{7}{8}\frac{\pi^2}{30} g T^4\,. \tag{6.13}$$

So, we have

$$\rho_b = \frac{8}{7}\rho_f = \frac{\pi^2}{30} g T^4\,, \tag{6.14}$$

where the scaling with the temperature again agrees with the simple dimensional analysis we performed in the last lecture.

Finally, for the pressure P we note that in the relativistic limit $p^2/E(p) = p = E(p)$, so that it trivially follows from the definitions in equations (6.3) and (6.4) that for bosons as well as fermions

$$P = \frac{1}{3}\rho = w\rho\,, \tag{6.15}$$

which nicely agrees with the equation of state parameter $w = \frac{1}{3}$ for radiation.

6.1.2 Non-relativistic particles

We can also analytically solve for n, ρ and P in the non-relativistic limit, i.e., for regular matter. In this case we have

$$E(p) = \sqrt{m^2 + p^2} \approx m + \frac{p^2}{2m}\,. \tag{6.16}$$

Let us define $x = p/\sqrt{2mT}$. Since the temperature is related to the average kinetic energy $T \sim \frac{p_{av}^2}{2m}$, which is much smaller than m, we find that $e^{E/T} \approx e^{m/T} \gg 1$. This means that the distribution function, as given in (6.1), is the same for bosons and fermions

$$f(p) = \frac{1}{e^{E(p)/T} \pm 1} \approx e^{-\frac{E}{T}} \approx e^{-\frac{m}{T}} e^{-x^2}\,. \tag{6.17}$$

This then gives for the number density

$$n = \frac{g}{(2\pi)^3} e^{-\frac{m}{T}} \int_0^\infty 4\pi (2mT)^{\frac{3}{2}} x^2 e^{-x^2} dx = \frac{g e^{-\frac{m}{T}} (2mT)^{\frac{3}{2}} \Gamma\left(\frac{3}{2}\right)}{4\pi^2} = g \left(\frac{mT}{2\pi}\right)^{\frac{3}{2}} e^{-\frac{m}{T}}\,, \tag{6.18}$$

where we used that $\Gamma(3/2) = \sqrt{\pi}/2$.

Worked problem 6.1: The number density for non-relativistic particles

Looking at the above expression, $n \propto e^{-\frac{m}{T}}$, explain the behavior of n for $T < m$.

Solution: Recall that T in our units with $k_B = 1$ is essentially the thermal energy, while m for $c = 1$ is the rest mass $E = mc^2$. Once the thermal energy drops below the rest mass energy particle-anti-particle pairs cannot be created anymore and they start to annihilate. This leads to an exponential decay of the particle density (as well as the anti-particle density).

In order to calculate the energy density, we use that $E \approx m$ and find to leading order from the definition in equation (6.3) that

$$\rho = mn = gm \left(\frac{mT}{2\pi}\right)^{\frac{3}{2}} e^{-\frac{m}{T}}\,. \tag{6.19}$$

Finally, we again calculate the pressure P as given in (6.4). Here we use that $p^2/E \approx p^2/m = (2mT)x^2/m$ and find, using the simplification leading to (6.18), that

$$P = \frac{g}{(2\pi)^3} \frac{e^{-\frac{m}{T}}}{3m} \int_0^\infty 4\pi (2mT)^{\frac{5}{2}} x^4 e^{-x^2} dx = \frac{g e^{-\frac{m}{T}} (2mT)^{\frac{5}{2}} \Gamma\left(\frac{5}{2}\right)}{12\pi^2 m}$$

$$= \quad gT\left(\frac{mT}{2\pi}\right)^{\frac{3}{2}} e^{-\frac{m}{T}} = nT\,, \tag{6.20}$$

where we used that $\Gamma(5/2) = 3\sqrt{\pi}/4$. Note that this is the familiar ideal gas law $P = nk_BT$ or after multiplying by the volume V: $PV = Nk_BT$.

Now as we argued above, the temperature T is much smaller than the mass m. This means that

$$P = nT = \frac{T}{m}\rho = w\rho \approx 0 \quad \text{for} \quad T \ll m\,. \tag{6.21}$$

So, we again reproduce our previous result that for non-relativistic matter we can neglect the pressure.

6.2 The effective number of relativistic species

The total radiation density is given by the sum over the contributions from all particles

$$\rho_r = \sum_i \rho_i = \frac{\pi^2}{30} g_\star(T) T^4\,, \tag{6.22}$$

where i runs over all standard model particles and $g_\star(T)$ is the effective number of degrees of freedom at temperature T, which we take to be the photon temperature. The sum over i can receive two contributions. One from relativistic particles that are in equilibrium with the photons, i.e., that have $T_i = T \gg m_i$. These contribute to $g_\star$ as follows

$$g_\star^{th}(T) = \sum_{i=bosons} g_i + \frac{7}{8} \sum_{i=fermions} g_i = g_b + \frac{7}{8} g_f\,, \tag{6.23}$$

where th stands for thermal equilibrium. However, particles can decouple so that they won't be in thermal equilibrium with the photons anymore. If these particles are relativistic, i.e., we have $T_i \neq T$ and $T_i \gg m_i$, then they contribute to $g_\star$

$$g_\star^{dec}(T) = \sum_{i=bosons} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{i=fermions} g_i \left(\frac{T_i}{T}\right)^4\,, \tag{6.24}$$

where dec stands for decoupled. We thus have

$$g_\star(T) = g_\star^{th}(T) + g_\star^{dec}(T)\,. \tag{6.25}$$

As we discussed in section 5, at $T \gg 100GeV$ all standard model particles are relativistic (cf. figure 31) and in thermal equilibrium with the photons (and each other).

So, let us calculate $g_\star$ for the standard model. We have the following contribution to g_b:

- The Z, W^+, W^- and the photon γ are all massless vectors and have two degrees of freedom each. Therefore, they contribute $4 \cdot 2 = 8$.
- Before the electroweak phase transition the Higgs scalar is a two vector whose entries are complex scalars so that it contributes $2 \cdot 2 = 4$.
- There are actually 8 gluons[15] that are all massless vectors, so that they contribute $8 \cdot 2 = 16$.

[15]This is not clear from figure 31 and follows from the more complicated nature of the strong force. The gluons are the entries in an $SU(3)$ matrix that has eight real independent entries.

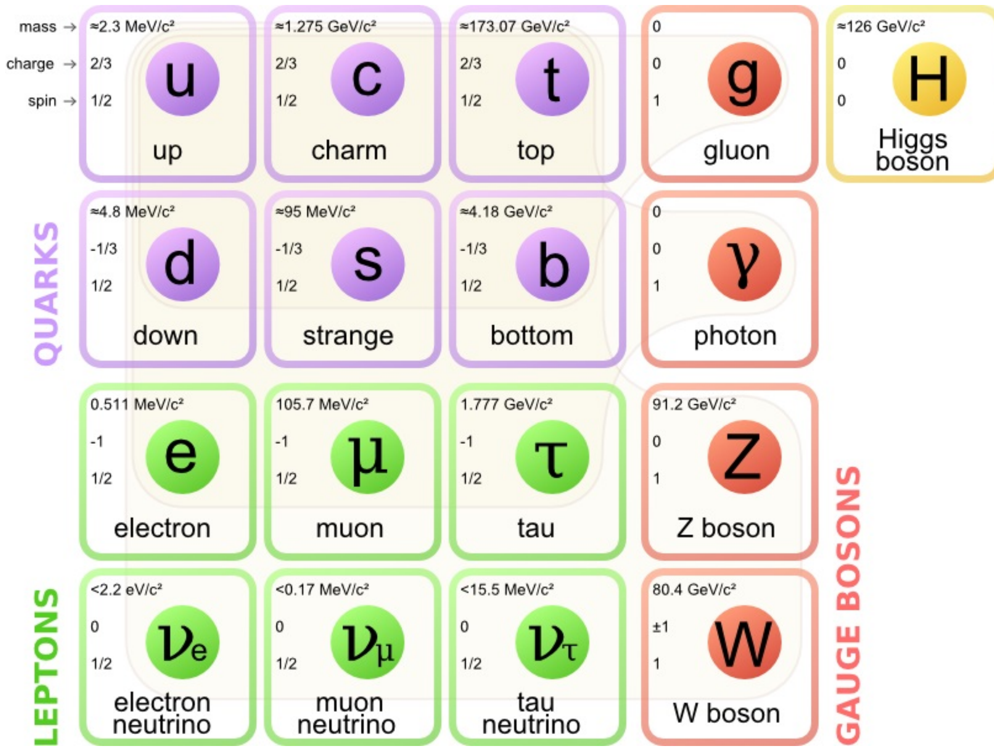


Figure 31: The known particles in our universe (taken from Wikipedia).

This leads to a total of $g_b = 28$.

Massive fermions have two possible spins and therefore have two internal degrees of freedom each. We take the neutrinos to be only left-handed, i.e., we assume that the right-handed neutrinos are very heavy and do not contribute. Fermions also have antiparticles that we need to include in our counting.[16] Then we find the following contributions to g_f:

- The left-handed neutrinos and their anti-particles contribute $3 \cdot 2 = 6$.
- The electron e, the μ and the τ contribute twice as much $3 \cdot 2 \cdot 2 = 12$.
- The six quarks can have three distinct charges under the strong force[17] which leads to an additional factor of 3 so that we have $6 \cdot 2 \cdot 2 \cdot 3 = 72$.

Thus, in total we have $g_f = 90$ and the value of $g_\star$ at temperatures well above a $100 GeV$ is

$$g_\star = g_b + \frac{7}{8} g_f = 28 + \frac{7}{8} 90 = 106.75\,. \tag{6.26}$$

In an expanding and cooling universe, particles will become non-relativistic. Before we discuss this in detail in the next subsection, let us mention the electroweak phase transition: At a temperature around $100 GeV$ the standard model of particle physics undergoes a transition during which the Higgs field develops a vacuum expectation value. This vacuum expectation value is actually what gives a mass to all the fields

[16] The bosons above are their own anti-particles so we did not need to include the anti-particles above.
[17] Each of them is a three vector on which the gluon $SU(3)$ matrix can act.

(particles) in the standard model. After this phase transition the $W^{\pm}$ and Z gauge fields have a mass. Massive vectors have three internal degrees of freedom so that this modifies our counting above. However, these new three degrees of freedom (one for each W^{+}, W^{-} and Z) come from the Higgs field that after this transition is only left with a single degree of freedom. So, the net number of degrees of freedom does not change during the electroweak phase transition.

6.3 Particle freeze-out

Once the temperature of the universe drops below the mass of a particle, the particle-anti-particle annihilation for this particle is favored compared to particle-anti-particle creation. This leads to an exponential decay of the particle number n, as derived in equation (6.18). This transition from relativistic to non-relativistic particle and the resulting annihilation of particles with their anti-particles is not instantaneous. Roughly 80% of the particles are annihilated in the interval $m > T > m/6$.

One effect of this so-called particle freeze out is, as we will discuss below, that the decrease in temperature of the universe is slowed down, since the particle-anti-particle annihilation deposits the energy contained in the annihilating particles into the remaining particles that are still in thermal equilibrium. But what happens to the particles themselves? Do they completely disappear?

In a non-expanding (but still somehow cooling) universe with vanishing chemical potential for these particles, this would be the case and the number density would keep decreasing exponentially with the temperature. However, as we mentioned above in subsection 5.3 (cf. equation (5.14)), in our universe there is a small baryon-anti-baryon asymmetry so that the particles cannot all annihilate since there aren't enough anti-particles around. This leads to the observed remaining baryons in our universe. Note, that in the standard model of particle physics heavy quarks (and leptons) decay to the lighter quarks (and leptons). For example, the top quark has an estimated lifetime of $5 \times 10^{-25}s$ so that any relic top quarks will quickly decay to up quarks. So, very quickly all massive particles (except the stable neutrinos) will decay to the three stable particles: the up quark, the down quark or the electron.

Another fate of non-relativistic particles in an expanding universe is that at a certain point their interaction rate Γ (which is proportional to their exponentially decaying number density n) becomes so small that it is smaller than the Hubble expansion H. In such a case the particles and anti-particles cannot find each other anymore and the annihilation stops. The exponential decay in the particle density followed by this so-called freeze-out is shown in the log-log-plot in figure 32. As we will see below in the next section, the neutrinos did undergo such a freeze-out in our early universe.

6.4 Evolution of the relativistic degrees of freedom

Having briefly discussed the potential fate of relativistic matter that becomes non-relativistic, let us return to the relativistic degrees of freedom of the standard model, during the time when our universe cools from a few hundred GeV to a few eV. The behavior of $g_{\star}(T)$ is shown in figure 33.

After the electroweak phase transition particles have their usual mass and the heaviest field, the top quark starts to become non-relativistic. This reduces $g_{\star}$ by

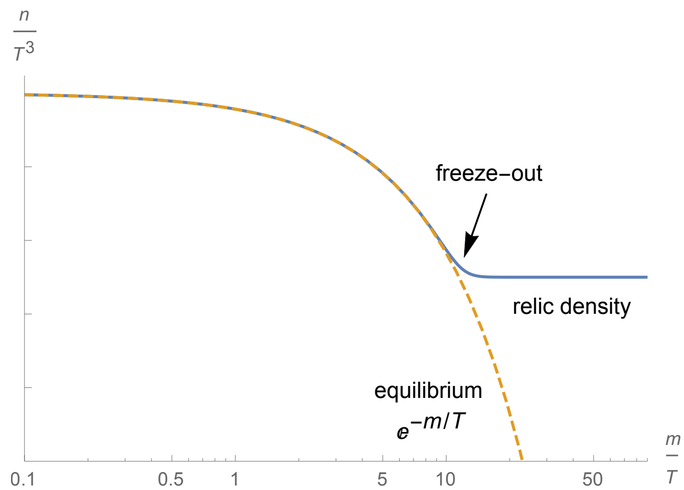


Figure 32: Once a particle becomes non-relativistic its number density decays exponentially. The Hubble expansion or a non-zero chemical potential can lead to a relic density.

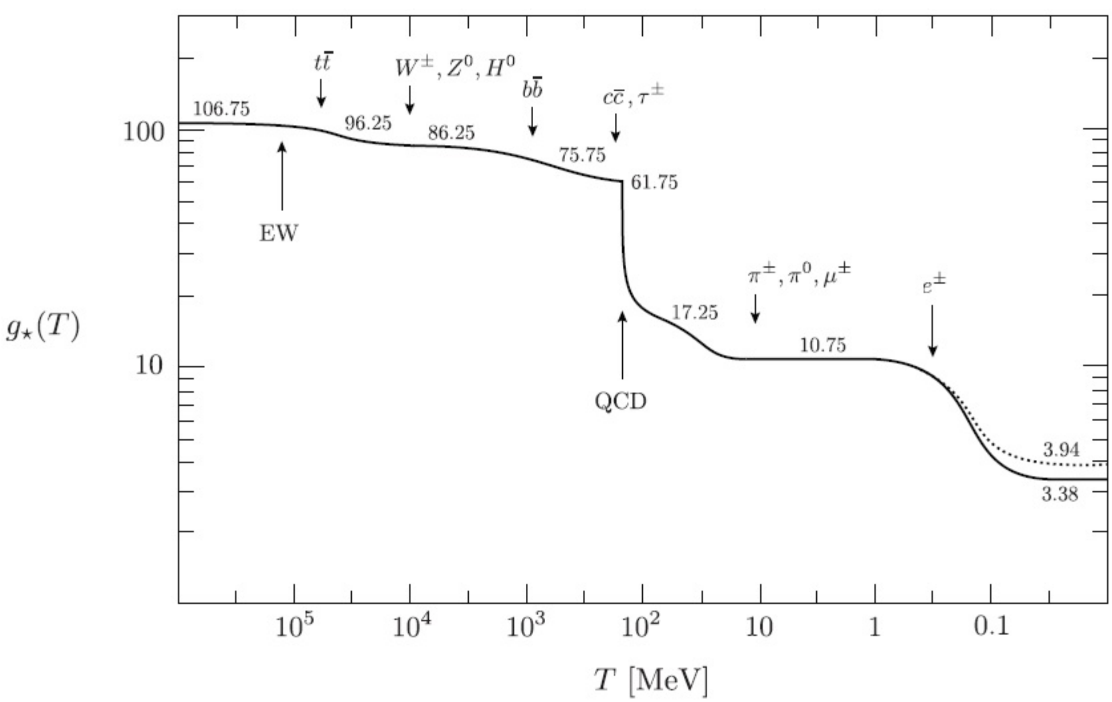


Figure 33: The evolution of the relativistic degrees of freedom in our early universe (taken from Daniel Baumann's "Cosmology" lectures).

$12 \cdot 7/8 = 10.5$. Next the massive vector bosons $W^{\pm}$ and Z and the Higgs scalar become non-relativistic, which reduces $g_\star$ by another $9 + 1 = 10$. After that the b and c quarks and the τ become also non-relativistic. At a temperature of roughly $150 MeV$ our universe undergoes another phase transition. The strong force becomes so strong that all quarks and gluons combine into uncharged bound states. For example, the u and d quarks combine into protons uud and neutrons udd. As we discussed above, all the particles that are combinations of three quarks are called baryons and even the lightest of them, the proton, has a mass of $1GeV \gg 150MeV$ so that after the QCD phase transition the baryons are all non-relativistic. However, there are also bound states of quarks and anti-quarks, the so-called mesons. The lightest mesons are the pi-mesons $\pi^+ = u\bar{d}$, $\pi^- = d\bar{u}$ and the π^0 which is a combination of $u\bar{u}$ and $d\bar{d}$. These three mesons have a mass of $135MeV$-$140MeV$ so that they will still be relativistic after the QCD transition. They become non-relativistic shortly before the μ leaving only the electron e, the photons and neutrinos as relativistic particles. As you can see from the graph something interesting is happening once the electrons become non-relativistic and we will discuss this in the next section.

Summary: The thermal universe

Using equilibrium thermodynamics, we derived the number density, energy density and pressure for relativistic and non-relativistic particles. This allowed us to rederive the previously obtained equation of state parameters $w = 1/3$ and $w = 0$.
Then we learned how to count relativistic degrees of freedom and how these evolve in our very early universe. We also learned of two ways how we can avoid ending up in a universe filled with only massless photons: 1) The previously discussed very small asymmetry between particles and anti-particles, corresponding to a non-zero chemical potential, leads to the remaining matter without anti-matter in our universe. 2) Another interesting possibility is the decoupling of certain particles: In an expanding universe, particles get diluted and if their coupling is sufficiently small then at a certain point they cannot annihilate with their anti-particles because they essentially cannot find each other anymore. This is what led to the neutrino cosmic background in our universe.

7 Neutrino cosmic background and dark matter

In the previous section we have seen that our universe roughly $10^{-14}s$ after the big bang was filled with a 'soup' of standard model particles that are all in thermal equilibrium. We described how during the expansion of our universe, due to the decrease in temperature, particles become non-relativistic. In this section we start at an energy of roughly $1MeV$, which corresponds to $1s$ after the big bang, and we will discuss the fate of electrons, positrons, photons and neutrinos. In order to do that it is useful to keep track of a conserved quantity, namely the entropy.

7.1 The entropy of our universe

In the second homework you used the first law of thermodynamics

$$TdS = dE + PdV \,, \tag{7.1}$$

to derive the continuity equation $\dot{\rho} + 3H(\rho + P) = 0$ under the assumption that the entropy is not changing in an expanding universe. The continuity equation follows from the Friedmann equations that we use to describe our universe which means that the entropy in our universe is conserved. This means it is a useful quantity to keep track of.

To do so we recall that the entropy and the energy are extensive quantities which satisfy

$$\partial_V E = \frac{E}{V} \,, \qquad \partial_V S = \frac{S}{V} \,. \tag{7.2}$$

Using this and $S = S(V, T)$ and $E(V, T)$ we find from equation (7.1)

$$\begin{aligned} T\partial_T S dT + T\partial_V S dV &= \partial_V E dV + \partial_T E dT + P dV \\ T\partial_T S dT + T\frac{S}{V} dV &= \frac{E}{V} dV + \partial_T E dT + P dV \,. \end{aligned} \tag{7.3}$$

Since dT and dV are independent, in particular the terms multiplying dV have to be equal, which gives

$$T\frac{S}{V} = \frac{E}{V} + P \quad \Rightarrow \quad S = \frac{E + PV}{T} \,. \tag{7.4}$$

We now define the entropy density

$$s \equiv \frac{S}{V} = \frac{\rho + P}{T} \,. \tag{7.5}$$

For our discussion of the early universe, we are interested in radiation with $P = \frac{1}{3}\rho$ for which the entropy density becomes

$$s = \frac{4}{3}\frac{\rho}{T} \,. \tag{7.6}$$

Note, that it easily follows from the scalings $\rho \sim a^{-4}$, $V \sim a^3$ and $T \sim a^{-1}$ that the entropy $S = s \cdot V$ is indeed constant.

In order to calculate the entropy density in the early universe, we have to sum over all particles, taking into account their energy density and temperature. We do this like in the previous section by defining

$$s = \frac{2\pi^2}{45} g_{\star S}(T) T^3 \,, \tag{7.7}$$

where we used equation (6.14) for the energy density ρ and T denotes the photon temperature. The quantity $g_{\star S}(T)$ again is the sum of the particles in thermal equilibrium with the photons and the decoupled particles $g_{\star S}(T) = g_{\star S}^{th} + g_{\star S}^{dec}(T)$. The effective number in thermal equilibrium is the same as before $g_{\star S}^{th} = g_{\star}^{th}$, where $g_{\star}^{th}$ was defined in equation (6.23). However, due to the different scaling with temperature, this is not true for the decoupled degrees of freedom. In particular, we have

$$g_{\star S}^{dec}(T) = \sum_{i=bosons} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8} \sum_{i=fermions} g_i \left(\frac{T_i}{T}\right)^3 \,, \tag{7.8}$$

which is different from $g_\star^{dec}(T)$ which has powers of 4 for the temperature (cf. equation (6.24)).

As we showed above, the entropy is preserved and in the case that no particles are created or destroyed then the particle density and entropy density scale in the same way, i.e., like $1/V \sim a^{-3}$. This means that their ratio n_i/s is constant. Now in the early universe in thermal equilibrium particles and anti-particles are constantly created and destroyed but as we discussed before the net baryon number cannot change due to perturbative standard model interactions. So, after baryogenesis the quantity $(n_b - n_{\bar{b}})/s$ is preserved and does not change during the evolution of our universe until today.

Another consequence of the fact that $sa^3 = const.$ is that, using equation (7.7), we find that

$$g_{\star S}(T)T^3a^3 = const. \quad \Rightarrow \quad T \propto g_{\star S}^{-\frac{1}{3}} a^{-1}\,. \tag{7.9}$$

This means that away from temperatures where particles become non-relativistic, we find that the factor of proportionality, i.e., the slope of the decrease of the temperature, is constant and depends on the relativistic degrees of freedom. If one species drops out of equilibrium because it becomes non-relativistic, then its entropy density (like its energy density) decays exponentially. However, the net entropy has to stay constant so the particle that becomes non-relativistic has to transfer its entropy to the particle species that are still in thermal equilibrium. For example, when electrons and positions are in thermal equilibrium, we have the reaction

$$e^- + e^+ \leftrightarrow \gamma + \gamma\,. \tag{7.10}$$

Once the temperature drops below the electron mass, the annihilation is strongly favored

$$e^- + e^+ \rightarrow \gamma + \gamma\,, \tag{7.11}$$

the electron and positron densities decay exponentially and their entropy is transferred to the photons. Since $g_{\star S}$ decreased by $\frac{7}{8}4$ during this decoupling of the electrons, the factor of proportionality between T and a^{-1} in equation (7.9) has increased. This is shown in figure 34.

7.2 Neutrino decoupling

In our universe things are more interesting and in particular the neutrinos play an interesting role. Via weak interactions like for example

$$e^- + e^+ \leftrightarrow \nu_e + \bar{\nu}_e\,, \tag{7.12}$$

they are kept in thermal equilibrium with the electrons. We have argued based on dimensional analysis that the cross-section for such interactions goes like $\sigma \sim \alpha^2/T^2$. However, this is not anymore true after the electroweak symmetry breaking. After this electroweak symmetry breaking the $W^\pm$ and Z bosons have a mass which provides a new scale and actually determines the weak interaction strength. A process with two initial and two final particles like the one in equation (7.12) scales like $\alpha^2/m_W^4 \approx 10^{-10}GeV^{-4}$. On dimensional grounds we then have

$$\sigma \sim \frac{\alpha^2}{m_W^4}T^2\,, \tag{7.13}$$

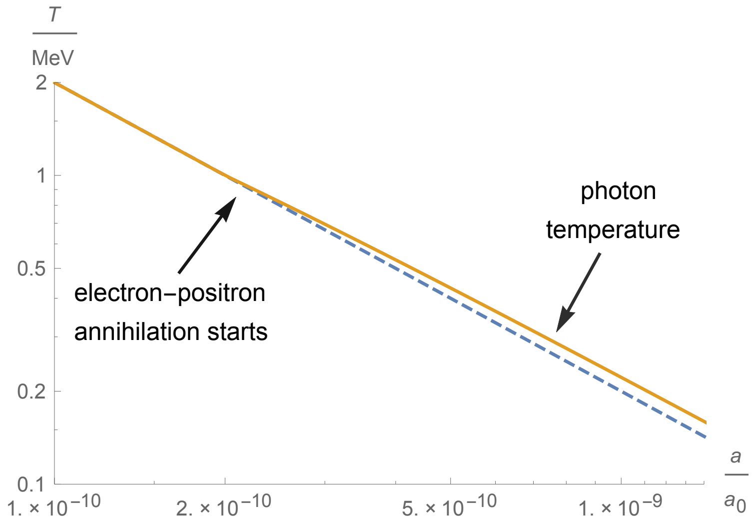


Figure 34: After the electrons and positrons become non-relativistic, they annihilate which changes the evolution of the temperature of the photons (orange line).

so instead of increasing with decreasing temperature, σ is now decreasing during the cooling of the universe.

Let us ask what this means for the ratio between the particle interaction rate $\Gamma = n\sigma v$ compared to the Hubble expansion rate $H \sim T^2/M_P$. Recalling that $n \sim T^3$ and $v \approx 1$, we get

$$\frac{\Gamma}{H} \sim \frac{\alpha^2}{m_W^4} T^5 \frac{M_P}{T^2} \approx \left(\frac{T}{MeV}\right)^3 . \tag{7.14}$$

What this estimate shows is that particles like the neutrinos that only interact via the weak force, will freeze-out at a temperature of roughly $1MeV$.

Worked problem 7.1: Neutrino mass and decoupling

Look up the current bounds on the mass of neutrinos. What does that mean for their decoupling?

Solution: The upper bound on the sum of the neutrino masses is $.12eV$. This means the heaviest neutrino has a mass that has to be below $.12eV$. This is much smaller than the freeze-out temperature. This means the neutrinos decouple from each other, while they are still relativistic. At a much later stage the neutrinos will becomes non-relativistic when the temperature drops below their (unknown) mass. Recall, that the CMB was emitted 380,000 years after the big bang, when the temperature was roughly $.3eV$. So, the neutrinos will become non-relativistic after the CMB was released.

7.3 The cosmic neutrino background

The only particles that are relativistic at such low energies are the electron, the photon and the neutrinos. The electron becomes non-relativistic at slightly lower temperature $T \lesssim m_e = .5MeV$. At this point the neutrinos are decoupled and the entropy of the electrons is transferred only to the thermal bath of the photons. This means that the neutrinos will have a lower temperature than the photons. In figure 34, the neutrino temperature would follow the dashed blue line, while the photon temperature corresponds to the orange line.

Let us make this more precise: Neglecting the decoupled neutrinos, we have before and after the electron decoupling

$$g_{\star S} = \begin{cases} 2 + \frac{7}{8}4 = \frac{11}{2} & T > m_e \\ 2 & T \ll m_e \end{cases} . \tag{7.15}$$

It then follows from equation (7.9) that the factor of proportionality between the photon temperature T_γ and a^{-1} changes due to the electron decoupling by a factor of $((11/2)/2)^{1/3} = (11/4)^{1/3}$. However, the same factor for the neutrinos does not change so that the neutrino temperature is slightly lower and given by

$$T_\nu = \left(\frac{4}{11}\right)^{\frac{1}{3}} T_\gamma \approx .71\, T_\gamma \,. \tag{7.16}$$

This means that our universe is filled in addition to the cosmic microwave background with a cosmic neutrino background that currently has a temperature of

$$T_{\nu,0} = \left(\frac{4}{11}\right)^{\frac{1}{3}} 2.725K \approx 1.95K \,. \tag{7.17}$$

We are currently not able to experimentally detect this cosmic neutrino background since the neutrinos have very low energies and interact only via the weak force. However, we have indirect evidence for their existence since their energy and entropy density affect the big bang nucleosynthesis and anisotropies in the cosmic microwave background, both of which we will discuss later in this course.

Next let us calculate the values of $g_\star$ and $g_{\star S}$ after the decoupling of the electrons. For $g_\star$ we have from equations (6.23)-(6.25)

$$g_\star(T) = g_\star^{th} + g_\star^{dec}(T) = 2 + \frac{7}{8} \cdot 2 \cdot 3 \cdot \left(\frac{4}{11}\right)^{\frac{4}{3}} \approx 3.36 \,. \tag{7.18}$$

Similarly, we find for $g_{\star S}$ from equation (7.8)

$$g_{\star S}(T) = g_{\star S}^{th} + g_{\star S}^{dec}(T) = 2 + \frac{7}{8} \cdot 2 \cdot 3 \cdot \frac{4}{11} \approx 3.91 \,. \tag{7.19}$$

These are the values shown in figure 35, that we have already seen above.[18] As is also clear from figure 35, before the neutrino freeze-out we have $g_\star = g_{\star S}$.

[18] The decoupling of the neutrinos is not instantaneous and not totally finished by the time the electrons become non-relativistic. This means that a small fraction of the electron entropy is transferred to the neutrinos. This is usually encoded in an effective number of neutrinos $N_{eff} \approx 3.046$ that replaces the 3 that counts the neutrinos in equations (7.18), (7.19) which leads to the values in figure 35.

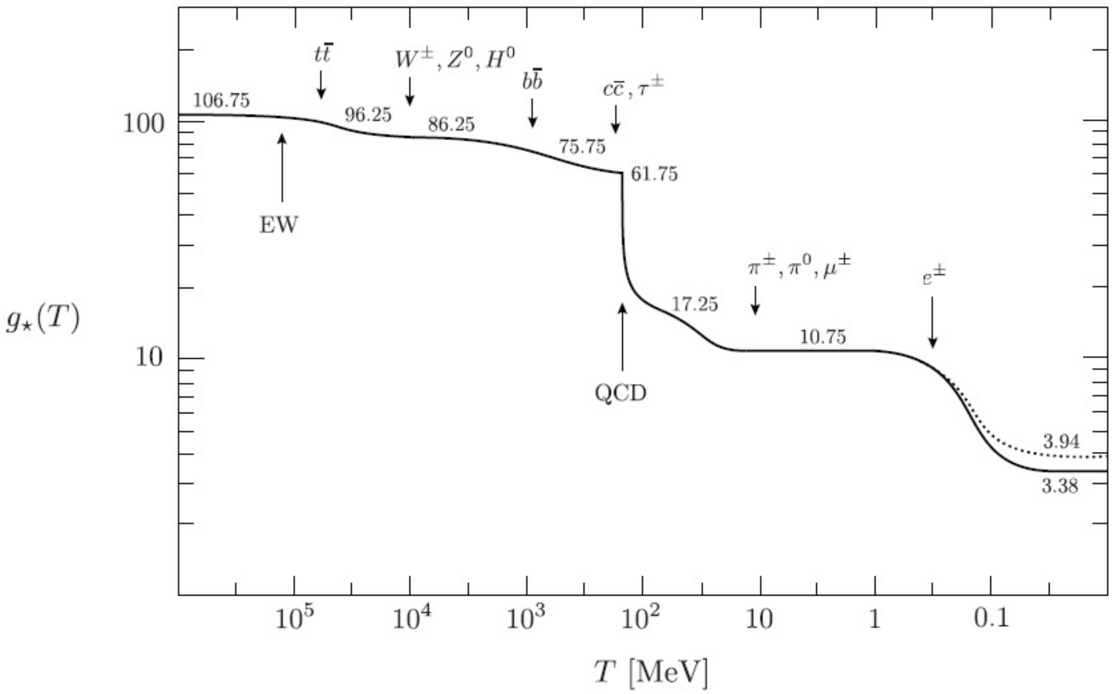


Figure 35: The evolution of the relativistic degrees of freedom in our early universe (taken from Daniel Baumann's "Cosmology" lectures).

7.4 Important events in our early universe

Before we start to discuss the deviation from equilibrium and big bang nucleosynthesis in section 8, it might be good to take stock and recall the important events of our universe in the time from $10^{-14}s$ to $380,000$ years after the big bang.

As we have discussed before the time of $10^{-14}s$ after the big bang corresponds to a temperature of $10TeV$ which is roughly the energy scale we can test in particle experiments like the LHC at CERN. We have also argued that the standard model particles are in thermal equilibrium at that time so that we have a hot and dense soup of standard model particles at this point. The expansion of the universe leads to a cooling of this plasma and the important events during the evolution and cooling of our universe are summarized in the table below.

Baryogenesis: As we discussed in subsection 5.3, there is an asymmetry between baryons and anti-baryons that cannot be explained by the standard model of particle physics. Thus, at energies above $1TeV$ there must be some new physics that generates this asymmetry. While there are many different theoretical ideas, there is no experimental test of any of these so we cannot associate a time to baryogenesis. Since the observed universe is neutral under the electric charge, there must be a similar asymmetry between electrons and positrons so that after their annihilation we are left with one electron for each proton.

Event	Time	Redshift	Temperature
Baryogenesis	?	?	?
EW phase transition	$2 \times 10^{-11}s$	10^{15}	$100GeV$
QCD phase transition	$2 \times 10^{-5}s$	10^{12}	$150MeV$
Neutrino decoupling	$1s$	6×10^9	$1MeV$
Electron-positron annihilation	$6s$	2×10^9	$500keV$
Big bang nucleosynthesis	$3min$	4×10^8	$100keV$
Matter-radiation equality	$6 \times 10^4 yrs$	3400	$.75eV$
Recombination	$2.6 - 3.8 \times 10^5 yrs$	1100-1400	$.26 - .33eV$
CMB	$3.8 \times 10^5 yrs$	1100	$.26eV$

Table 1: A summary of important events in our early universe.

Electroweak-phase transition: During this phase transition that we discussed in the previous section, the particles get their mass due to the so called Higgs effect. Once the standard model particles are massive, they start to drop out of equilibrium whenever the temperature of the universe (i.e., the thermal bath) becomes smaller than their mass. Then the particles start to annihilate with their anti-particles and their number densities decrease exponentially. The remaining matter in our observed universe is due to the matter-anti-matter asymmetry mentioned above.

QCD phase transition: The strong force is weaker at higher energies (temperatures) and becomes stronger and stronger during the cooling of the universe. Around $150MeV$ the strong force is so strong that free gluons and quarks cannot exist anymore and all the quarks are bound into so called baryons and mesons. These are bound states that are neutral under the strong force. The lightest baryons are the familiar proton and neutron. There are also heavier baryons and mesons that can be lighter than the proton and neutron but all of these are unstable and quickly decay. So, a little bit after the QCD phase transition we are left with essentially only protons and neutrons that are the building blocks for the atomic nuclei.

Neutrino decoupling: As we discussed today, at around $1MeV$ the weak interaction becomes so weak that particles that are only charged under the weak force, i.e., the neutrinos, decouple from the thermal plasma. These neutrinos, similarly to the photons in the CMB, give rise to a cosmic neutrino background that is slightly colder than the CMB and is difficult to observe directly. At the time of decoupling the three neutrinos are still relativistic and during the cooling of the universe they become non-relativistic whenever the temperature is smaller than their respective mass.

Electron-positron annihilation: Around $T \sim m_e \approx 511keV$ the electrons and positrons become non-relativistic and transfer their energy and entropy into the photons only (since the neutrinos are decoupled already). This slows down the decrease in the temperature of the photons a little bit so that the photons today have a temperature that is a little bit larger than the temperature of the cosmic neutrino background.

Big bang nucleosynthesis: One of the greatest successes of the big bang cosmology is that it correctly predicts the observed abundance of elements in our universe. We will discuss in the next section 8 how the protons and neutrons in our universe combine into atomic nuclei. Using nuclear physics, we can predict the ratios of numbers of different elements in the early universe and these predictions agree with what we observe. Any kind of new physics that can appear beyond the standard model is severely constraint by this success.

Recombination: Once the average energy of the photons drops below $.33eV$ the tail of high energy photons is sufficiently small to allow for neutral atoms to form. This process in which electrons and protons combine[19] takes roughly 100,000 years and at its end the universe is filled with clouds of neutral atoms and the cosmic microwave background.

The cosmic microwave background: Once the electrons and nuclei combine into neutral atoms, the photons can stream freely until today. The observation of this cosmic microwave background does not only tell us about the universe 380,000 years after the big bang but the incredible homogeneity of the CMB also strongly motivates a short phase of accelerated expansion in our very early universe, the so-called inflation. The small deviations from homogeneity in the CMB photons we observe together with their polarization provide detailed information about this period of inflation.

7.5 Dark Matter

As we have seen in subsection 3.2, roughly 80% of the matter in our universe is not in the form of standard model particles but rather in the form of dark matter that consists of one (or multiple) unknown particle species. Since we don't know what these particles are and how they interact, we cannot say for sure how their observed energy density arises. However, if we assume that the dark matter and standard model particles are in thermal equilibrium in the early universe, then the evolution of the dark matter particles should be describable with the tools we developed so far. In particular, since the dark matter is not visible and so difficult to detect, it can at most interact with the standard model particles via the weak force (or via an unknown even weaker force). If we assume that the mass of the dark matter particles is above their decoupling scale, then they would become non-relativistic before they decouple from the standard model particles and their number density starts to exponentially decay. Then at a certain point their interaction rate becomes so small that $\Gamma \sim H$ which leads to a freeze-out and a relic density of dark matter particles. This scenario can lead to the observed

[19] Don't ask why this is called *re*combination. The electrons and nuclei were never combined before that point.

amount of dark matter and experiments have already substantially constraint the cross-section and mass of such weakly interacting massive particles (often called WIMPs). This is shown in figure 36 where the constraints on the WIMP mass is plotted against the WIMP-nucleon cross section. The upper right region is excluded by experiments.

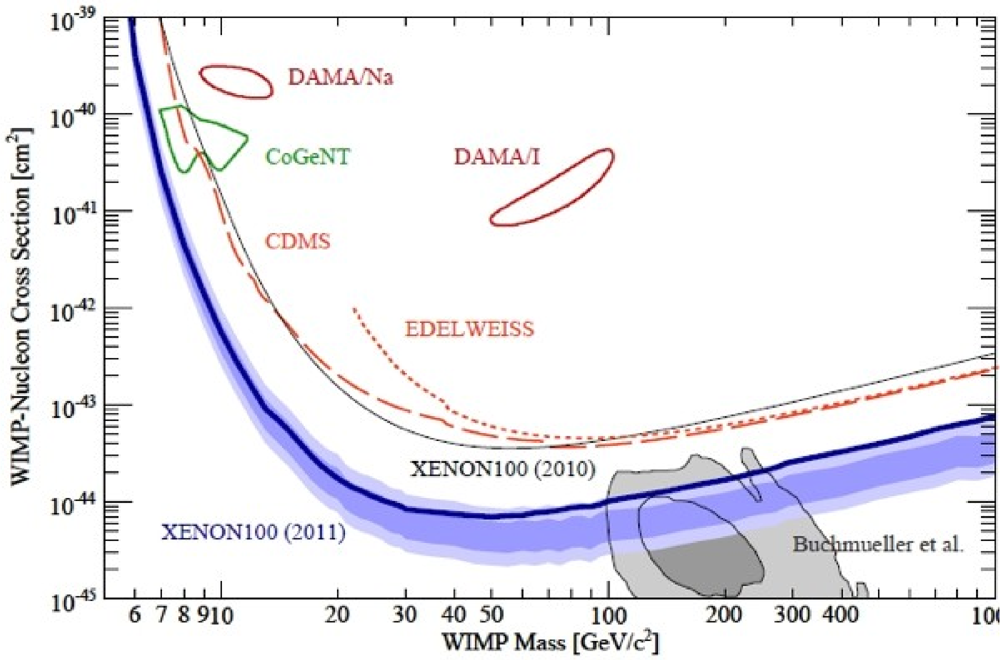


Figure 36: Some experimental bounds on weakly interacting massive particles.

While some experiments have claimed a detection of dark matter in the past, other experiments could not reproduce these findings. So, we have to wait for future experiments to shed light on the nature of the dark matter in our universe.

Summary: The cosmic neutrino background and dark matter

We have learned about entropy and entropy density and how they evolve in our early universe. Then we used that to understand how the electron-positron-annihilation in our early universe slowed down the cooling of the photons. The neutrinos on the other hand had decoupled already 1s after the big bang and therefore cooled faster than the photons, leading to a slightly lower temperature for them $T_\nu \approx .71 T_\gamma$.
Subsection 7.4 provides an intermediate brief summary of all important events discussed in part II of these notes. This leaves out dark matter because our current understanding of it is very limited. In particular, the once popular idea of weakly interacting massive particles (WIMPs) has been tested in the relevant mass range and experiments have come up empty handed. However, there are many other theoretical ideas for dark matter, some of which will get tested in the future.

8 Big bang nucleosynthesis

In this section we discuss the Boltzmann equation that allows one to describe the evolution of processes in our universe that are not in equilibrium. Then we discuss the formation of light elements during big bang nucleosynthesis and the recombination of electrons and protons into neutral hydrogen.

8.1 The Boltzmann Equation

The number density in the absence of interactions (or in equilibrium) scales like the inverse volume, i.e., like a^{-3}, since it is a density. This means that it satisfies the equation

$$0 = \frac{1}{a^3}\frac{d(na^3)}{dt} = \frac{dn}{dt} + 3\frac{\dot{a}}{a}n = \frac{dn}{dt} + 3Hn\,. \tag{8.1}$$

As we discussed in the previous section, there are ample interactions in which two particles interact and become two new particles. These can be schematically written as

$$1 + 2 \leftrightarrow 3 + 4\,, \tag{8.2}$$

which means that particle 1 and 2 annihilate and become particles 3 and 4 (and vice versa). Such interactions together with decays of single particles are the most relevant processes in the early universe since the interaction of three or more particles is much more unlikely because these three or more particles would have to be all very close at the same time.

The Boltzmann equation describes the evolution of the number density n_1 of for example particle 1 in the presence of interactions. Here we focus on the interaction (8.2), in which case the Boltzmann equation is given by

$$\frac{1}{a^3}\frac{d(n_1 a^3)}{dt} = -\langle\sigma v\rangle n_1 n_2 + c\, n_3 n_4\,, \tag{8.3}$$

where the first term describes the reduction of n_1 due to annihilation of particles 1 with 2, while the second term describes the production of 1 particles (and 2 particles) due to the annihilation of 3 and 4 particles. The free parameter c can be related to the thermally averaged cross-section $\langle\sigma v\rangle$: We know from equation 8.1 that the right-hand-side of equation 8.3 has to vanish in thermal equilibrium, i.e., for $n_i = n_i^{eq}$. This gives

$$c = \frac{n_1^{eq} n_2^{eq}}{n_3^{eq} n_4^{eq}}\langle\sigma v\rangle\,. \tag{8.4}$$

The Boltzmann equation then becomes

$$\frac{1}{a^3}\frac{d(n_1 a^3)}{dt} = -\langle\sigma v\rangle\left(n_1 n_2 - \frac{n_1^{eq} n_2^{eq}}{n_3^{eq} n_4^{eq}}\, n_3 n_4\right)\,. \tag{8.5}$$

This can be rewritten as

$$\frac{d\log(n_1 a^3)}{d\log(a)} = -\frac{\Gamma_1}{H}\left(1 - \frac{n_1^{eq} n_2^{eq}}{n_3^{eq} n_4^{eq}}\,\frac{n_3 n_4}{n_1 n_2}\right)\,, \tag{8.6}$$

where $\Gamma_1 = n_2\langle\sigma v\rangle$. The above equation determines the evolution of the number density for particles species 1 as a function of $a(t)$. Since $a(t)$ grows with t in our universe we can essentially think of the above Boltzmann equation as determining the evolution of species 1 with time. We see that Γ_1/H plays a crucial role in determining the evolution of $n_1 a^3$. If the interaction rate Γ_1 becomes small compared to the Hubble rate H, we have a freeze out and the number density of n_1 scales like a constant times a^{-3}.

To describe the evolution of all the particles in our early universe one has to solve simultaneously all the corresponding *coupled* Boltzmann equations. This is of course only possible numerically and goes beyond what we will discuss in class. Here we will focus on a few simple interesting cases that we can discuss more or less analytically and using the equilibrium results from the previous lectures. We will henceforth drop the superscript *eq* and just write n_i for the number densities in equilibrium.

Chemical potentials

Before we discuss big bang nucleosynthesis it is useful to review the effect of a non-zero chemical potential μ. In the phase space distribution function (see for example equation (6.1)) a non-zero chemical potential leads to

$$f_{\pm}(p) = \frac{1}{e^{(E(p)-\mu)/T} \pm 1}\,, \tag{8.7}$$

where $\mu(T)$ is generically a complicated function. While again each particle can have a different chemical potential, chemical equilibrium, which is reached via interactions, leads to relations between the chemical potentials. For example, interaction like the ones in equation (8.2) lead to

$$\mu_1 + \mu_2 = \mu_3 + \mu_4. \tag{8.8}$$

Non-zero chemical potentials will modify the expression for, for example, the number density, so that for non-relativistic particles in equilibrium it is given by

$$n = g \left(\frac{mT}{2\pi}\right)^{\frac{3}{2}} e^{\frac{\mu-m}{T}}\,. \tag{8.9}$$

However, if we take ratios of number densities in which the chemical potential cancels due to equation (8.8), then we don't really need the values of the chemical potentials.

Worked problem 8.1: Photon chemical potential

Photons can interact with electrons via a double Compton scattering

$$e^- + \gamma \leftrightarrow e^- + \gamma + \gamma\,. \tag{8.10}$$

What does that mean for the chemical potential of photons?

Solution: In equilibrium the chemical potentials on the left-hand-side and the right-hand-side have to be equal, which means

$$\mu_e + \mu_\gamma = \mu_e + 2\mu_\gamma\,. \tag{8.11}$$

This means that photons have vanishing chemical potential

$$\boxed{\mu_\gamma = 0}\,. \tag{8.12}$$

8.2 Big bang nucleosynthesis

Big bang nucleosynthesis refers to the formation of atomic nuclei during the cooling of our early universe. Recall that after the QCD phase transition around $150MeV$ quarks form colorless bound states that include protons and neutrons. During the continuous cooling of our universe, the number densities of these non-relativistic baryons is exponentially decaying until, due to the initial asymmetry between baryons and anti-baryons, we are left with a residual amount of baryonic matter in the form of protons and neutrons and heavier nuclei. The protons and neutrons can bind via the strong force into atomic nuclei and via the weak force neutrons and protons can convert into each other. All these processes are initially in equilibrium and we want to understand with which relic abundance of nuclei we are left, once these processes drop out of equilibrium due to the cooling of our universe.

The two reasons why we can actually do that without solving many coupled Boltzmann equations are firstly that essentially no elements heavier than Helium are created during big bang nucleosynthesis, so that we can just focus our attention on Hydrogen and Helium and secondly that initially we have only neutrons and protons in equilibrium without any relevant number of heavier nuclei.

Worked problem 8.2: The beginning of nucleosynthesis

We have seen that deuterium starts to form around $T \approx .1MeV$. What is roughly the corresponding time for the beginning of big bang nucleosynthesis?

Solution: We can cheat slightly and look-up in table 1 that $T = 1MeV$ corresponds to $1s$ after the big bang. Since we are in the radiation dominated phase, we have $T \propto a(t)^{-1} \propto t^{-\frac{1}{2}}$. A reduction in temperature by a factor of ten hence leads to an increase in time by 100. So, big bang nucleosynthesis starts roughly after $100s \sim 1-2$ minutes. It doesn't take very long and all the primordial helium in our universe formed within minutes.

8.2.1 Protons and neutrons

At temperatures above $1MeV$ protons and neutrons are in equilibrium due to weak interactions of the form

$$n + \nu_e \leftrightarrow p^+ + e^-\,. \tag{8.13}$$

The chemical potential is the average energy needed to add an extra particle ("$dE = \mu dN$"). Electrons and neutrinos are much lighter than neutrons and protons and the particles are non-relativistic so that $E \sim m$. We therefore conclude that the chemical potentials for electrons and neutrinos are negligible small, so that equation (8.8) tells us that $\mu_p = \mu_n$. Taking the ratio of the proton and neutron number densities, the chemical potential then simply cancels (see equation (8.9)) and we find

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{\frac{3}{2}} e^{-\frac{m_n - m_p}{T}} . \tag{8.14}$$

Recalling the proton and neutron masses $m_p = 938.27MeV$ and $m_n = 939.57MeV$, we see that their ratio is very close to 1 and their difference is $m_n - m_p = 1.3MeV$. So at large temperatures $T \gg 1MeV$ we have the same number of neutrons and protons, while at energies below $T \sim 1MeV$, the ratio of the neutron to proton number density is exponentially decaying. However, as we have seen above when we discussed neutrinos in subsection 7.2, processes that involve the weak interactions like the one in equation (8.13) will become irrelevant at energies below roughly $1MeV$, since $\Gamma/H \approx 1$ for $T \approx 1MeV$ (see equation (7.14)). Actually, a more careful analysis reveals that the weak interactions become irrelevant at $T \approx .8MeV$ which leads to

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{\frac{3}{2}} e^{-\frac{m_n - m_p}{T}} \approx e^{-\frac{1.3MeV}{.8MeV}} \approx .2 . \tag{8.15}$$

Once the temperature drops further the finite lifetime of the neutron becomes important. In particular, a free neutron can decay via

$$n \to p^+ + e^- + \bar{\nu}_e , \tag{8.16}$$

which leads to an exponential decay of the neutron number density

$$\frac{n_n}{n_p} \to \frac{n_n}{n_p} e^{-\frac{t}{886s}} \approx .2 e^{-\frac{t}{886s}} , \tag{8.17}$$

where we used that the mean lifetime of a free neutron is $886s$. The decay of the neutrons stops once they are bound into nuclei which happens around $t \approx 330s$ which leads to

$$\left.\frac{n_n}{n_p}\right|_{t \approx 330s} \approx .14 . \tag{8.18}$$

8.2.2 Heavier nuclei

Let us study a process that involves the production of the lightest nucleus that is not just a proton, i.e., deuterium. One neutron and one proton can form deuterium (and a photon):

$$n + p^+ \leftrightarrow D^+ + \gamma . \tag{8.19}$$

As we argued above, the photon's chemical potential vanishes so that the chemical potentials cancel in the following ratios

$$\frac{n_D}{n_n n_p} = \frac{3}{4} \left(\frac{2\pi}{T} \frac{m_D}{m_n m_p}\right)^{\frac{3}{2}} e^{-\frac{m_D - m_n - m_p}{T}} , \tag{8.20}$$

where we used $g_n = g_p = 2$ and $g_D = 3$.[20] The ratio between the masses is approximately $2/m_p$, however, the difference in the mass of the deuterium and its two constitutions is the binding energy $m_n + m_p - m_D \approx 2.2MeV$. At energies well below the proton and neutron masses, i.e., at $T \ll 1GeV$, the number densities of protons and neutrons are not exponentially decaying anymore but are determined by the non-zero baryon number in our universe, i.e., by equation (5.14)

$$n_p \sim n_n \sim n_b \sim 10^{-9} n_\gamma = 10^{-9} \frac{2\zeta(3)}{\pi^2} T^3 , \tag{8.21}$$

where we used equation (6.11) for the photon number density. Using this in equation (8.20), we get

$$\frac{n_D}{n_p} \approx 8 \left(\frac{T}{m_p} \right)^{\frac{3}{2}} e^{\frac{2.2MeV}{T}} 10^{-9} . \tag{8.22}$$

This implies that for $T = 1MeV$, we have $n_D/n_p \approx 10^{-12}$ and for roughly $T \approx .066MeV$ we have $n_D/n_p \approx 1$. This means that at temperatures above $T \approx .1MeV$ the deuterium abundance is negligible and the same is true for even heavier nuclei.

8.2.3 Nucleosynthesis

Now we have all pieces in place and can discuss the creation of nuclei that are not just a proton. Our starting point are protons and neutrons. As we mentioned before, processes involving more than two particles are very rare so that the initial process must be the formation of deuterium from one proton and one neutron as shown in equation (8.19). Only once deuterium is formed, which as we saw above happens around $T \approx .066MeV$, can Helium be produced via

$$\begin{aligned} D + p^+ &\leftrightarrow {}^3He + \gamma , \\ D + {}^3He &\leftrightarrow {}^4He + p^+ . \end{aligned} \tag{8.23}$$

The binding energy of ${}^4He$, B_{He}, is larger than that of deuterium B_D. This leads to an enhancement of the number density of Helium compared to that of deuterium

$$\frac{n_{He}}{n_D} \propto e^{\frac{B_{He} - B_D}{T}} . \tag{8.24}$$

This is similar to equation (8.22), where deuterium is favored at low temperatures, except that here we don't have a suppression factor. This means that helium is almost immediately produced after deuterium and that all neutrons end up in ^{4}He nuclei. Since each ^{4}He atom contains two neutrons, this allows us to easily determine the fraction of helium to hydrogen in our universe

$$\frac{n_{He}}{n_H} = \frac{n_{He}}{n_p} \approx \frac{\frac{1}{2} n_n}{n_p} = 7\% . \tag{8.25}$$

This answer is very close to a full numerical analysis that solves all the coupled Boltzmann equations and which gives something like $6.2\% \approx \frac{1}{16}$. Since the mass of a Helium nucleus is roughly four times as large as the proton mass, we find that roughly one fourth of the mass of ordinary matter in our early universe is in the form of Helium and the rest in the form of Hydrogen. This perfectly agrees with observations and is one of the great successes of big bang nucleosynthesis and shown in figure 37.

[20] Deuterium is the spin 1 combination of the proton and neutron, so $g_D = 2s + 1 = 3$. The corresponding spin 0 particle is unstable.

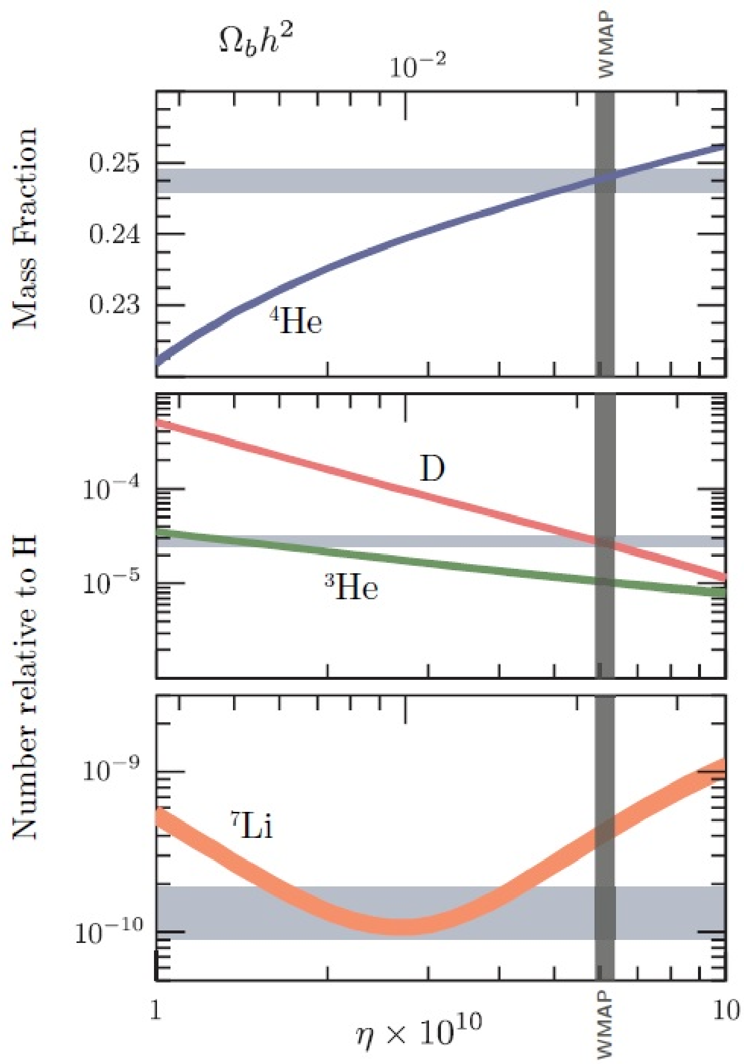


Figure 37: The mass fraction of several light nuclei as theoretical predicted (colored band) and the observational constraints (grey bands). The value of the top horizontal axis is fixed by $\Omega_b h^2$ as given above in equation (3.9). This translates into $\eta = n_b/n_\gamma$ on the bottom horizontal axis (cf. equation (5.14)).

Beyond Helium You probably wonder why heavier atomic nuclei don't form during big bang nucleosynthesis (and how they appeared in our universe). The reason that they aren't formed from protons, neutrons, deuterium and helium is the following: As we have seen above, before helium can be formed, protons and neutrons need to first combine to form a substantial amount of deuterium. During this time the universe keeps cooling and the nuclei loose part of their kinetic energy, which makes it harder to overcome the Coulomb barrier (i.e., to bring together two positively charge nuclei). More importantly, once a large amount of ^4He is formed, these can only combine to form ^8Be which is unstable and decays faster than it can be formed. Very small amounts of Tritium and ^3He that are also created during big bang nucleosynthesis can combine with ^4He to form ^7Li of which we observe tiny amounts today.[21] So big bang nucleosynthesis produces only very light elements. As briefly mentioned before, the heavier elements that we see today in our universe and that we are made of are created in the first stars through nuclear fusions. Once the heaviest of these stars explode in supernovae, these heavy elements are released into the interstellar medium and can become part of second-generation stars and form planets.

The abundance of the different nuclei in our early universe as a function of time is shown in figure 38.

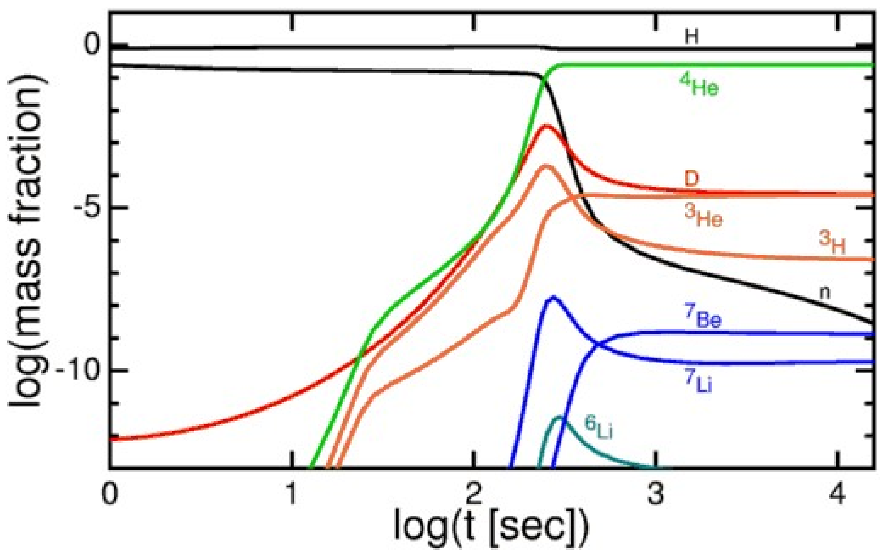


Figure 38: The mass fraction of several light nuclei as a function of time. The logarithm base is 10 and nucleosynthesis essentially happens around 200 seconds, i.e., around 3 minutes after the big bang. (Figure taken from `http://www.astro.ucla.edu/~wright/BBNS.html`).

8.3 Recombination and photon decoupling

After nuclei are formed around $T \approx .06MeV$, which corresponds roughly to a time of 3 minutes after the big bang, our universe contains a soup of positively charged nuclei, free electrons and photons (as well as decoupled neutrinos). During the further expansion and cooling of our universe the energy density of radiation decays like a^{-4}, while the energy density in the non-relativistic matter decays like a^{-3}. Roughly 60,000 years after the big bang the energy densities in radiation and matter are equal and our universe enters its matter dominated era. Another 200,000 years later electrons and nuclei start to form neutral atoms, during a period that is usually called recombination.[22] Once the recombination ends and the universe essentially consists of neutral atoms, the photons in the cosmic microwave background can stream freely until today

[21] Actually we observe slightly more Lithium than theoretical predicted (see figure 37) which might require small modifications of big bang nucleosynthesis.

[22] A more accurate name would be *combination* since they have never been combined before.

and tell us about the universe 380,000 years after the big bang as well as much earlier times.

8.3.1 The Saha equation

The process that keeps electrons, protons and photons in equilibrium during the first 200,000 years after the big bang is

$$e^- + p^+ \leftrightarrow H + \gamma \,. \tag{8.26}$$

At a temperature of $T \approx 1eV$ all particles except the photons are non-relativistic and they all are in (chemical and thermal) equilibrium due to the above process. Recalling that the photon chemical potential vanishes, we can look at the following ratio in which the chemical potentials cancel[23]

$$\frac{n_H}{n_p n_e} = \left(\frac{m_H}{m_e m_p}\frac{2\pi}{T}\right)^{\frac{3}{2}} e^{\frac{m_p+m_e-m_H}{T}} \,, \tag{8.27}$$

where we used that $g_H = 4 = g_e g_p$. Using that $m_p \approx m_H$, that the binding energy of hydrogen is $m_p + m_e - m_H = 13.6eV$ and the fact that our universe is electrically neutral which implies $n_e = n_p$, we find

$$\frac{n_H}{n_e^2} = \left(\frac{2\pi}{m_e T}\right)^{\frac{3}{2}} e^{\frac{13.6eV}{T}} \,. \tag{8.28}$$

Next, we define the free electron fraction as the ratio of free electrons to baryons

$$X_e \equiv \frac{n_e}{n_b} \,. \tag{8.29}$$

As we have seen above in equation (8.25), more than 90% of the baryon number is due to the protons, that can be in the form of positively charged nuclei n_p or in the form of neutral hydrogen n_H, so that

$$n_b \approx n_p + n_H = n_e + n_H \approx 10^{-9}\frac{2\zeta(3)}{\pi^2}T^3 \,, \tag{8.30}$$

where we used equation (8.21). From the definition in equation (8.29) we then find

$$\frac{1-X_e}{X_e^2} = \frac{(n_p+n_H)-n_e}{n_e^2}(n_p+n_H) = \frac{n_H}{n_e^2}n_b \,. \tag{8.31}$$

Using this in equation (8.28) we find the Saha equation

$$\boxed{\frac{1-X_e}{X_e^2} = 10^{-9}\frac{2\zeta(3)}{\pi^2}\left(\frac{2\pi T}{m_e}\right)^{\frac{3}{2}} e^{\frac{13.6eV}{T}} \,.} \tag{8.32}$$

[23]Here we use n_H to denote the neutral hydrogen only so that $n_p \neq n_H$.

8.3.2 Recombination

The Saha equation allows us to get an estimate for the energies at which recombination happened. Taking the onset of recombination as the temperature $T_{beginning}$ at which $X_e = .9$ and the end of recombination as the temperature T_{end} at which $X_e = .1$, we find from equation (8.32) that $T_{beginning} \approx .35eV$ and $T_{end} \approx .30eV$. The reason that these results are so much smaller than the $13.6eV$ binding energy is that there are many, many more photons than baryons and that the black body spectrum of the photons has a tail of high energy photons that keep the Hydrogen ionized until the average temperature of the photon bath is well below the binding energy of Hydrogen.[24]

8.3.3 Photon decoupling

The so-called time of last scattering at which the electrons and photons scatter for the last time via Thompson scattering

$$e^- + \gamma \leftrightarrow e^- + \gamma \,, \tag{8.33}$$

is actually happening even later around a time when $X_e \approx .01$. We see this as follows: The cross-section for Thompson scattering is $\sigma_T \approx 2 \times 10^{-3} MeV^{-2}$ and the corresponding interaction rate is given by

$$\Gamma_T \approx n_e \sigma_T = n_b X_e \sigma_T \approx 10^{-9} \frac{2\zeta(3)}{\pi^2} T^3 X_e \sigma_T \,. \tag{8.34}$$

In order to determine the temperature at decoupling T_{dec} we have to check when the above interaction rate is of the same size as the Hubble expansion rate. During matter domination the Hubble rate is given by [25]

$$H = H_0 \left(\frac{T}{T_0} \right)^{\frac{3}{2}} \,. \tag{8.35}$$

This implies

$$\Gamma_T(T_D) = H(T_D) \qquad \Leftrightarrow \qquad T_D^{\frac{3}{2}} X_e(T_D) = 10^9 \frac{\pi^2}{2\zeta(3)} \frac{H_0}{\sigma_T \, T_0^{\frac{3}{2}}} \,. \tag{8.36}$$

We can numerically solve this equation and find $T_D \approx .26eV$ and $X_e(T_D) \approx .003$. The temperature $T_D \approx .26eV$ corresponds to a time of 380,000 years after the big bang and a redshift of $z \approx 1100$.

[24] The binding energy, i.e., the ionization energy for the first electron in ^4He is the highest of any atom in the periodic table, namely $24.6eV$. Therefore, Helium nuclei combine with electrons before the protons.

[25] This follows from the first Friedmann equation

$$H^2 = \frac{8\pi G}{3} \rho_m \propto \left(\frac{a_0}{a(t)} \right)^3$$

and the fact that $a(t) \propto 1/T$.

Summary: Nucleosynthesis

Arguably the biggest success of the big bang model is the predictions of element abundances in our very early universe. The initial soup of particles cools and eventually starts forming nuclei. Their abundance can be calculated using standard nuclear physics and we found that, surprisingly, essentially only Helium is being formed. The ratio of Helium nuclei and protons (i.e., Hydrogen nuclei) perfectly matches observations. This so-called nucleosynthesis happened during the first few minutes of our universe.

After more than 200,000 years the universe has cooled sufficiently so that electrons can combine with the nuclei to form neutral atoms. It takes until 380,000 years after the big bang to have a sufficiently large number of neutral atoms that allows a free streaming of the cosmic microwave background that we observe today.

Part III - Inflation

9 Inflation solves early universe problems

Having discussed the thermal history of our universe and in particular its evolution at times larger than 10^{-14} seconds after the big bang, we will now venture even closer to the initial singularity and discuss the theory of inflation. We will first layout the original problems that cosmologists were facing before Guth, Linde, Albrecht and Steinhardt invented inflation in the 1980's. Then we explain how inflation solves these problems.

9.1 Beyond ΛCDM

To describe the evolution of our universe so far, we have been able to use well tested particle and nuclear physics and two extra ingredients: the cosmological constant Λ and cold dark matter (CDM), where cold refers to the fact that this matter behaves like non-relativistic matter with equation of state parameter $w = 0$. This so called ΛCDM model seems to correctly describe the evolution of our universe from 10^{-14} seconds after the big bang until today. However, we have already seen that there has to be something else that we don't understand yet: The asymmetry between matter and anti-matter requires processes in the earlier universe that go beyond the standard model of particle physics. In addition, it seems that our universe underwent a period of inflation during very early times. While experiments provide us with ever improving bounds on different inflationary models, they have not yet singled out one particular model of inflation so we will discuss a variety of different models and their features. Since the theory of inflation is less well tested, we should ask what its generic predictions are and which of those we observe. As discussed above, cosmologists up until 1980 were faced with some problems that get resolved, if our very early universe underwent a period of inflation. However, the absence of these problems is then not a prediction but rather a post-diction of inflation since it was invented to resolve these issues. Does inflation make in addition any generic predictions that we can test? Yes, as we discuss at the end of this course, all inflationary models predict a nearly scale invariant spectrum for the density perturbation that are tiny deviations from an entirely homogeneous universe. The imprint of these density perturbations has been observed in the cosmic microwave background. So, it is fair to say that the observational evidences for the theory inflation are pretty robust and our universe most likely underwent such a phase at a time that could be as early as 10^{-34} seconds after the big bang.

9.1.1 The horizon problem

The first problem in an early universe that is only dominated by radiation and matter is called the horizon problem. Recall that the photons of the cosmic microwave background decoupled 380,000 years after the big bang and they constitute the best black body spectrum ever observed in nature. This black body spectrum has the same temperature everywhere in the sky. In particular this means that all the photons that come to us from one side have the same temperature as the photons that come from the opposite side. This seems only possible, if these photons have been in causal contact with each other so that they can be in thermal equilibrium. We have previously discussed that there are abundant interactions in the early universe that establish a local equilibrium but we haven't discussed the size of these local patches in equilibrium. In order to do so, it is useful to work again with the conformal time τ.

The metric in conformal times is given by (see equation (4.3) above)

$$ds^2 = a(\tau)^2 \left(-d\tau^2 + \frac{dr^2}{1 - Kr^2} + r^2 d\Omega \right) , \tag{9.1}$$

and a radially traveling light ray satisfies

$$ds = 0 \qquad \Rightarrow \qquad d\tau = \frac{dr}{\sqrt{1 - Kr^2}} \equiv d\chi \, . \tag{9.2}$$

In particular this means that in the (χ, τ)-plane light rays travel along straight lines at $45°$ angles. For each point P in the (χ, τ)-plane we can draw a future light-cone and a past light-cone by drawing two straight lines at $\pm 45°$ angles that intersect in P. Every point in spacetime inside the past light-cone can send information to the point P and every point in the future light-cone can receive information from the point P. We have already discussed these causally connected parts in subsection 4.1 above. The radius of the future light-cone is the event horizon and the radius of the past light-cone is the particle horizon (see figure 39).

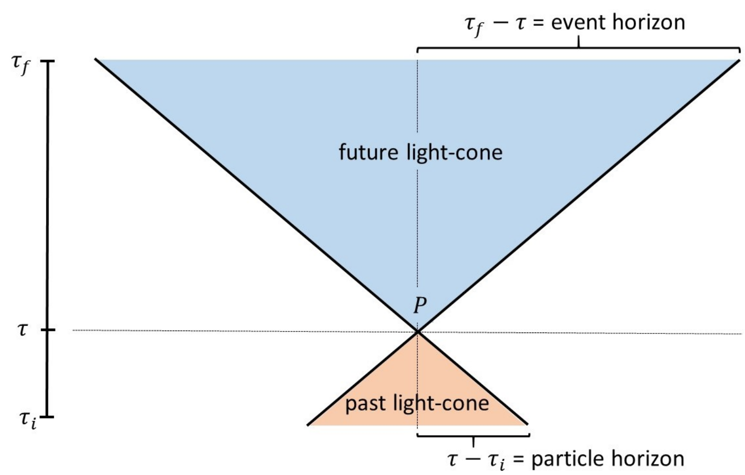


Figure 39: The future and past light-cone associated to the point P.

The question we have to ask is whether all the points on the surface of last scattering that we observe today have been in causal contact or not? The answer is no! This means that in an early universe that is radiation dominated the photons that are coming from one side of the universe and are reaching us today have never been in causal contact with the photons coming from the other side of the universe. So, why are they both having the same temperature? This is the horizon problem and is depicted in figure 40.

The initial singularity with $a(\tau = 0) = 0$ has non-zero (and maybe even infinite) comoving spatial size. If we assume that the points on this initial surface are not

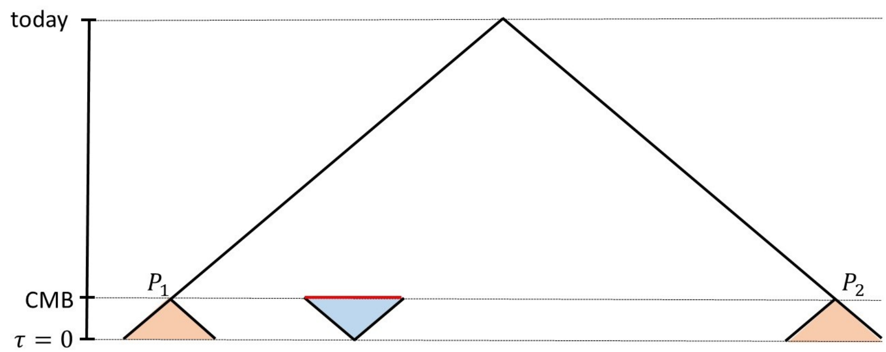


Figure 40: The point P_1 and P_2 have past light-cones that do not intersect so these two points have never been in causal contact. The blue cone is the future light-cone of one point on the initial singular surface and the red line shows the maximal surface size that could have been in causal contact.

particularly fine-tuned to have all the same initial conditions then we expect thermal equilibrium to lead to the same temperature only for rather small patches of the sky. A simple calculation for our universe reveals that only photons coming from within a 2° angle should be in thermal equilibrium. This means that the surface of last scattering should consist of roughly $4\pi/(2^\circ * 2\pi/360^\circ)^2 \approx 10^4$ different patches that have not been in thermal equilibrium.

We will see later that there are actually small fluctuations in the temperature of the CMB of the order of .01%. While this is incredibly small, it raises the question of how different causally disconnected patches would have to be in order not to worry about the fine tuning of the initial surface. A natural expectation would be a change of temperature between different regions of the order of $1K$, i.e., of the order of $T_{CMB,0} = 2.725K$. Additionally, once we know about these small fluctuations in the CMB then there is actually an even better posed horizon problem: The fluctuations in the temperature are correlated on scales much larger than 2°, so how is this possible, if these regions have not been in causal contact?

9.1.2 The flatness problem

The second problem is also related to fine tuning. Recall from subsection 3.2 that the curvature contribution to the energy density of our universe is very small. This means that the normalized total energy density today is $\Omega_T \approx 1$. More specifically, from equations (3.5) and (3.2), we have

$$\Omega_{tot} - 1 = \frac{K}{\dot{a}(t)^2} = \frac{3K}{8\pi G\, \rho_c(t)\, a(t)^2}\,. \tag{9.3}$$

Today the experimental bound is (see equation (3.6))

$$\frac{K}{\dot{a}(t_0)^2} < .005\,. \tag{9.4}$$

Let us look at the last term in equation (9.3). If the energy density is dominated by a cosmological constant then $\rho_c(t)$ is constant, however, during a matter dominated era we have $\rho_c(t) \propto a(t)^{-3}$ and during the radiation dominated era we have $\rho_c(t) \propto a(t)^{-4}$. This means that any small deviation of Ω_{tot} from one will grow during the radiation and matter dominated era. Likewise, going back in time we find that the value of Ω_{tot} must have been incredibly close to 1 in the early universe. At matter-radiation equality ($z = 3400$) we have roughly

$$\Omega_{tot} - 1 < .005 \frac{1}{3400} \approx 1.5 \times 10^{-6} \,, \tag{9.5}$$

and at the time of the electro-weak phase transition with $T \approx 100 GeV$ and $z = 10^{15}$ we have

$$\Omega_{tot} - 1 < .005 \frac{1}{3400} \left(\frac{3400}{10^{15}} \right)^2 \approx 1.7 \times 10^{-29} \,. \tag{9.6}$$

Finally, at the Planck scale $T \approx M_P = 2.2 \times 10^{18} GeV$ we would have

$$\Omega_{tot} - 1 < .005 \frac{1}{3400} \left(\frac{3400}{10^{15} \cdot 2.2 \times 10^{16}} \right)^2 \approx 3.5 \times 10^{-62} \,. \tag{9.7}$$

This seems like an incredible fine tuning, which can't be explained by just $K = 0$. Even in a flat universe we would expect some locally changing value of K, since the spacetime is dynamical in general relativity. So why is the value in our (visible) universe so incredibly fine-tuned?

9.1.3 The monopole problem

Alan Guth was thinking about so-called grand unified theories (GUT) in which all forces of the standard model of particle physics are unified in a single force around an energy scale of approximately $10^{16} GeV$. This single force is broken into the strong, weak and electromagnetic force at energies below $10^{16} GeV$ and usually this phase transition leads to unwanted relics like for example magnetic monopoles. The existence of such heavy particles that would be produced during the phase transition can overclose the universe ($\Omega \gg 1$). So, Guth's original motivation for inflation was to remove the overabundance of these heavy relics in GUT theories.

9.2 Inflation

Inflation is a period of accelerated expansion of our universe that happened at very early times. Here we will focus on the case in which the universe is approximately exponentially expanding as is the case during an era that is dominated by a cosmological constant. Such a period solves the three problems above, if it lasts sufficiently long. We can make this more precise and determine a minimal amount of exponential expansion that is required to solve each of the problems above.

9.2.1 Solving the horizon problem

If in the very early universe there would have been a phase during which $a(t)$ grows exponentially, then a small patch in local thermal equilibrium could be stretched to the size of the surface of last scattering that we observe today and thereby solve the horizon problem.[26] More precisely, if the early universe would have been in a phase with $a(t) \propto e^{Ht}$, then the beginning of the universe would not be at $t = 0$ anymore but at $t = -\infty$. This would likewise push the initial conformal time τ_i to $-\infty$, while any finite period of inflation pushes τ_i to a negative but finite value. This can then allow for causal contact of all the points we observe on the surface of last scattering as is shown in figure 41.

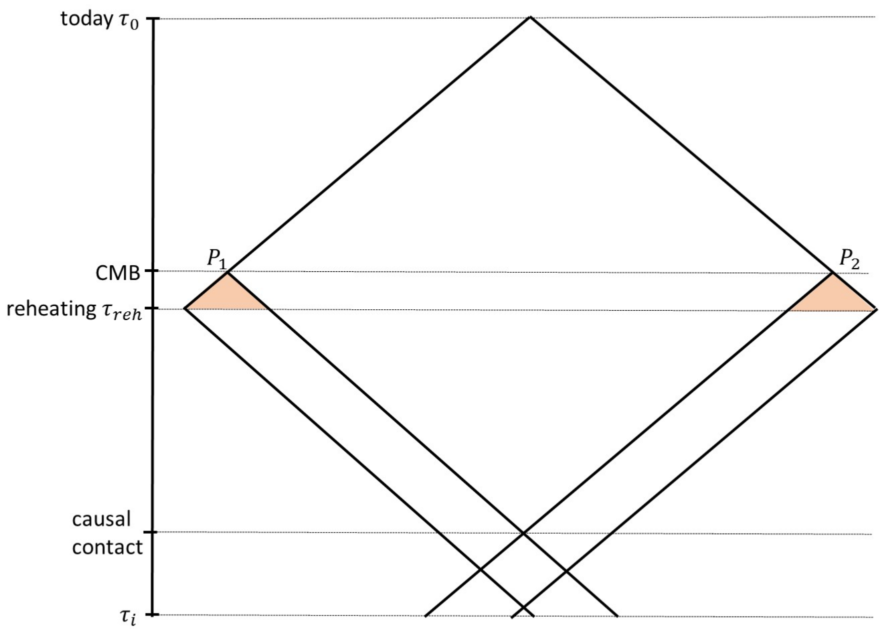


Figure 41: The point P_1 and P_2 have past light-cones that can intersect, if we have a period of exponential expansion that pushes τ_i sufficiently far back.

Before we discuss how such a period of inflation can arise in our very early universe, we would like to get an estimate for how long it would have to last in order to solve the horizon problem. From the diagram 41, it is clear that we need the conformal time between the beginning of inflation τ_i and the end of inflation which is around the reheating time τ_{reh} to be at least as large as the time between reheating and today at τ_0.

Recalling that $d\tau \equiv a(t)^{-1}dt$ we get

$$\tau_{reh} - \tau_i = \int_{\tau_i}^{\tau_{reh}} d\tau' = \int \frac{dt'}{a(t')} = \int \frac{da}{a\dot{a}} = \int_{a_i}^{a_{reh}} \frac{da}{a^2 H_{inf}} \approx \frac{1}{a_i H_{inf}} - \frac{1}{a_{reh} H_{inf}} \approx \frac{1}{a_i H_{inf}}\,, \tag{9.8}$$

where we used that the Hubble constant during inflation is approximately constant and that $a_i \ll a_{reh}$ due to the exponential expansion during inflation.

To get a very simple (but fairly accurate estimate) we assume that the universe after the end of inflation is radiation dominated so that $a(t) = a_0\sqrt{t/t_0}$ and $a^2 H =$

[26] Actually things are more complicated: Inflation itself needs somewhat special initial conditions to start, so that the horizon problem is substantially alleviated but not completely solved. After inflation ends the temperature of the universe is essentially zero, because $T \propto a(t)^{-1}$ and $a(t)$ grows exponentially during inflation. However, the inflaton field carries energy that is then used to homogeneously reheat the universe.

$const. \approx a_{reh}^2 H_{inf}$ since $a(t)$ is a smooth function. This then leads to

$$\tau_0 - \tau_{reh} = \int_{\tau_{reh}}^{\tau_0} d\tau' = \int_{a_{reh}}^{a_0} \frac{da}{a^2 H} \approx \frac{1}{a_{reh}^2 H_{inf}}(a_0 - a_{reh}) \approx \frac{a_0}{a_{reh}^2 H_{inf}}\,, \tag{9.9}$$

where we used that $a_0 \gg a_{reh}$. We then find the constraint

$$\begin{aligned} \tau_{reh} - \tau_i &\gtrsim \tau_0 - \tau_{reh} \\ \frac{1}{a_i H_{inf}} &\gtrsim \frac{a_0}{a_{reh}^2 H_{inf}} \\ \frac{a_{reh}}{a_i} &\gtrsim \frac{a_0}{a_{reh}} \approx \frac{T_{reh}}{T_0}\,, \end{aligned} \tag{9.10}$$

where we used in the last line that the temperature scales like the inverse of the scale factor. For example, for inflation with an energy scale of $T_{reh} \approx 10^{14} GeV$ we then find

$$\frac{a_{reh}}{a_i} \gtrsim 10^{26} \approx e^{60} \equiv e^{N_e}\,. \tag{9.11}$$

We see that the expansion factor during inflation is gigantic so that one defines the number of e-folds N_e that is given by the logarithm to basis e of the expansion factor. The number of e-folds required to solve the horizon problem depends on the energy scale that we chose to be $10^{14} GeV$ above and that is not known. For the inflationary models that are currently being tested the energy scalar is very high $10^{15} - 10^{16} GeV$ and the standard values in the literature are $N_e = 50 - 60$ which nicely matches with the value derived above.

Worked problem 9.1: The duration of inflation

From figure 41 above, we see that we roughly need to double the age of the universe in conformal time, since τ_0 corresponds to today with $t_0 = 13.8 Gyr$ and `CMB` correspond to 380,000 yrs. Let us determine how long an additional period of inflation has to last in real time: From equation (9.11) we see that the universe needs to grow by e^{60}. Determine the duration of inflation that gives rise to such 60 e-folds for $H_{inf} = 10^{14} GeV$ and $H_{inf} = 100 TeV$.

Solution: During inflation we have an exponential growth $a(t) \approx e^{H_{inf} t}$, so that we find that the time needed for 60 e-folds is simply

$$t = \frac{60}{H_{inf}} = \frac{60\hbar}{H_{inf}} \approx \frac{4 \times 10^{-14} eV \cdot s}{H_{inf}}\,. \tag{9.12}$$

For $H_{inf} = 10^{14} GeV$ we find $t = 4 \times 10^{-37} s$ and for $H_{inf} = 100 TeV$ we get $t = 4 \times 10^{-28} s$. So, the required minimal amount of inflation that leads to 60 e-folds happens in an incredibly short interval in real time.

9.2.2 Solving the flatness problem

The flatness problem is solved because during inflation $a(t) \propto \dot{a}(t)$. The large grows of $\dot{a}(t)$ then suppresses the term $\frac{K}{\dot{a}(t)^2}$ in equation (9.3). If we would take $\frac{K}{\dot{a}(t)^2}$ to be initially some order one number and we want it to be sufficiently small at the end of inflation to explain the observed value, then we need for example for inflation ending slightly below the GUT scale around $10^{14}GeV$ that

$$\frac{K}{\dot{a}(t_i)^2} \approx 1\,, \tag{9.13}$$

$$\Omega_{tot}(t_{reh}) - 1 = \frac{K}{\dot{a}(t_{reh})^2} \lesssim .005\frac{1}{3400}\left(\frac{3400}{10^{15}\cdot 10^{12}}\right)^2 \approx 1.7\times 10^{-53}\,, \tag{9.14}$$

$$\Rightarrow \quad \frac{\dot{a}(t_{reh})}{\dot{a}(t_i)} \approx \frac{a(t_{reh})}{a(t_i)} \gtrsim \left(1.7\times 10^{-53}\right)^{-\frac{1}{2}} \approx e^{60}\,, \tag{9.15}$$

where we used that during inflation $a(t) \approx e^{H_{inf}t}$ with H_{inf} the constant Hubble parameter during inflation. We see that we again need roughly 60 e-folds of inflation to solve the flatness problem and this value is of course related to the energy scale at which we would like to solve the flatness problem.

Worked problem 9.2: Energy dependence of the number of e-folds

If inflation happens at higher energy scales, we usually need more e-folds to solve the flatness (and also the horizon problem). How many e-folds do we require to solve the flatness problem for an energy of $100TeV$?

Solution: We simply repeat the above calculation for $100TeV = 10^5GeV$ instead of $10^{14}GeV$ and find

$$\frac{K}{\dot{a}(t_i)^2} \approx 1\,,$$

$$\Omega_{tot}(t_{reh}) - 1 = \frac{K}{\dot{a}(t_{reh})^2} \lesssim .005\frac{1}{3400}\left(\frac{3400}{10^{15}\cdot 10^{3}}\right)^2 \approx 1.7\times 10^{-35}\,,$$

$$\Rightarrow \quad \frac{\dot{a}(t_{reh})}{\dot{a}(t_i)} \approx \frac{a(t_{reh})}{a(t_i)} \gtrsim \left(1.7\times 10^{-35}\right)^{-\frac{1}{2}} \approx e^{40}\,. \tag{9.16}$$

So, in order to solve the flatness problem with such low scale inflation we would only need 40 e-folds of inflation instead of 60.

9.2.3 Solving the monopole problem

It is intuitively clear that a period of inflation will also solve the monopole problem, provided that the monopoles and other relics are not produced after inflation. If magnetic monopoles would be produced at the GUT scale, then a period of inflation that takes places at lower energies dilutes away all the relics and leaves us with an almost empty universe. The amount of inflation necessary to solve the monopole problem is usually a little bit lower and roughly 30 e-folds are enough to sufficiently dilute the relics so that they don't have any impact on the cosmological evolution and wouldn't be observable today. However, if there is a GUT theory at energies around $10^{16}GeV$ then the reheating temperature after inflation has to be lower than this $10^{16}GeV$. The current upper bound on the energy scale during inflation is around the GUT scale so that the reheating temperature is also bounded from above by the GUT scale and there is no tension between grand unified theories and inflation.

9.3 A period of inflation from a scalar field

A great problem with a very early period in our universe that is dominated by a large cosmological constant is that the energy density of the cosmological constant does not decay during the expansion, while it does so for matter and radiation. This means, as we discussed previously, that a large cosmological constant is inconsistent with our observed universe. So, what we need for inflation is a fluid that mimics a large cosmological constant for a short period of time and then transfers its energy into the other particles of our universe during a reheating process and afterwards essentially disappears. This can be accomplished by a scalar field, i.e., a spinless particle similar to the Higgs field (but most likely not the standard model Higgs particle).

The action for such a scalar field ϕ, that is called the inflaton, is given by

$$S = \int d^4x\sqrt{-g}\left(-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\right) . \tag{9.17}$$

Here $V(\phi)$ is the potential for the scalar field that we leave arbitrary. You can think of this scalar field as a 'ball' rolling in a potential. The only difference here is that ϕ in principle depends not only on the time t but also on the spatial coordinates x^i so the value of ϕ can change throughout space. You are probably familiar with this from electrodynamics where the electric and magnetic fields can vary through space and time.

As derived in appendix B, the variation of the above action with respect to the metric leads to the following energy density and pressure for a scalar field in the FRW universe

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}\frac{(\nabla\phi)^2}{a^2} + V(\phi)\,, \tag{9.18}$$

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - \frac{1}{6}\frac{(\nabla\phi)^2}{a^2} - V(\phi)\,. \tag{9.19}$$

Note, that if the spatial variation $\nabla\phi$ and the time variations $\dot{\phi}$ vanish, then we have $\rho_\phi = V(\phi) = -P_\phi$ so that the scalar field behaves exactly like a cosmological constant! Guth's original idea was that the scalar field sits at a false minimum of the potential as shown in figure 42. The scalar field will then lead to an effective large cosmological constant and a period of inflation. After quantum tunneling through the barrier the scalar field will roll down the potential to the true minimum and again lead to an effective cosmological constant that could be the observed value of the cosmological constant today if $V_{today} \approx 10^{-120}M_P^4$.

The problem with Guth's original proposal is that the quantum tunneling will happen via the nucleation of spatial bubbles. Inside these bubbles the field is on its way to the true vacuum and outside of the bubbles is the false vacuum. Since the energy

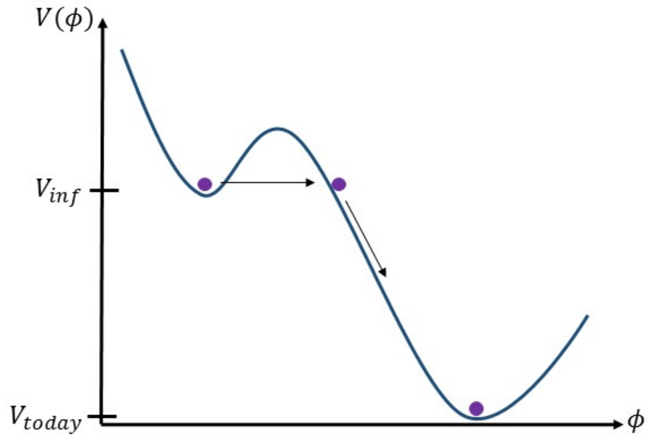


Figure 42: The scalar field is initially trapped in a false vacuum leading to a period of inflation with a large cosmological constant proportional to V_{inf}. After quantum tunneling the scalar field rolls to the true minimum at which the value of the potential is tiny $V_{today} \sim 10^{-120} M_P^4$.

is smaller inside of these bubbles, the bubble walls will expand outwards. However, it turned out that these bubbles cannot be large enough to contain our entire universe and the collision of multiple bubbles would lead to inhomogeneities that are larger than the ones we observe. So, there is no nice way of ending inflation in this case.

However, shortly after Guth's original idea, Linde, Albrecht and Steinhardt proposed another kind of inflation in which the scalar field is not trapped in a false vacuum but rather rolls very slowly since the scalar potential is very flat as is shown in figure 43. When the inflaton reaches a steeper part of the potential, its kinetic energy becomes important and it does not behave like a cosmological constant anymore and inflation ends. When the scalar field reaches its true minimum, it will oscillate and couplings to the standard model particles can transfer the kinetic energy of the inflaton into standard model particles which leads to a reheating of the universe and the start of our hot universe that we can describe so well with thermodynamics.

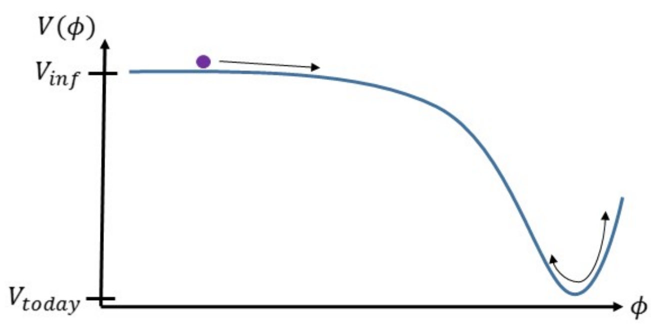


Figure 43: The potential is very flat so that the scalar field rolls very slowly leading to a period of inflation with a large cosmological constant proportional to V_{inf}. Once the potential steepens, inflation ends and the scalar field rolls to its true minimum with $V_{today} \sim 10^{-120} M_P^4$.

Summary: Problems in the early universe

We have seen in this section that our very early universe is highly fine-tuned. Light from the CMB, arriving at the earth from opposite directions, has exactly the same temperature although the light originates from places that should have never been in causal contact. A period of inflation, during which our universe grows in size by potentially a factor of 10^{23} cubed, can solve this and other problems like the flatness problem. Such a period of inflation would need to last only for a split second and require the existence of a new particle called the inflaton. Such a particle can give rise to an exponentially expanding universe during the very early universe as well as today, i.e., it could also be the underlying explanation behind the dark energy that we observe in our current universe.

10 Slow-roll inflation

In the previous section we showed how a period of inflation can solve several problems that we encounter in our very early universe. In this section we are studying the relevant equations for inflation as well as a few exemplary models.

10.1 A scalar field

As we have seen in the previous section, a new scalar field, called the inflaton, can lead to a temporary phase of inflation. To make this precise we vary the action for the scalar field

$$S = \int d^4x\sqrt{-g}\left(-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\right)\,, \tag{10.1}$$

with respect to the scalar field to derive its equation of motion. In order to do this recall that the FRW metric takes the form

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2+a(t)^2\gamma_{ij}dx^i dx^j \equiv -dt^2+a(t)^2\left(dx_i^2 + K\frac{x_i^2 dx_i^2}{1-Kx_i^2}\right)\,. \tag{10.2}$$

This then leads to

$$\begin{aligned}
\delta S &= \int d^4x a(t)^3\left(-g^{\mu\nu}\partial_\mu\phi\partial_\nu\delta\phi - V'(\phi)\delta\phi\right)\\
&= \int d^4x\left[\partial_\nu\left(a(t)^3 g^{\mu\nu}\partial_\mu\phi\right) - a(t)^3V'(\phi)\right]\delta\phi\\
&= \int d^4x\left[-\partial_t\left(a(t)^3\partial_t\phi\right) + \partial_i\left(a(t)\gamma^{ij}\partial_j\phi\right) - a(t)^3V'(\phi)\right]\delta\phi\\
&= \int d^4x\left[-3\dot{a}(t)a(t)^2\dot{\phi} - a(t)^3\ddot{\phi} + a(t)\nabla^2\phi - a(t)^3V'(\phi)\right]\delta\phi\\
&= \int d^4x\left(-a(t)^3\right)\left[\ddot{\phi} + 3\frac{\dot{a}(t)}{a(t)}\dot{\phi} - \frac{\nabla^2\phi}{a(t)^2} + V'(\phi)\right]\delta\phi\,,
\end{aligned} \tag{10.3}$$

where we used the short-hand notation $V'(\phi) = \partial_\phi V(\phi)$. So, the equation of motion for a scalar field in an FRW universe is given by

$$\ddot{\phi} + 3H\dot{\phi} - \frac{\nabla^2\phi}{a(t)^2} + V'(\phi) = 0\,. \tag{10.4}$$

Once inflation starts, $a(t)$ grows exponentially so that the term with the spatial derivatives of ϕ quickly becomes unimportant and the above equation reduces to

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0\,. \tag{10.5}$$

Worked problem 10.1: Scalar field EOM from continuity equation

Derive the above scalar field equation of motion in equation (10.5) from the continuity equation. Use that

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi)\,, \tag{10.6}$$

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi)\,. \tag{10.7}$$

Solution: We simply plug in the above into the continuity equation (cf. equation (1.34))

$$\begin{aligned} \dot{\rho}_\phi + 3H(\rho_\phi + P_\phi) &= 0 \\ \dot{\phi}\ddot{\phi} + \dot{\phi}V'(\phi) + 3H\dot{\phi}^2 &= 0 \\ \ddot{\phi} + 3H\dot{\phi} + V'(\phi) &= 0\,. \end{aligned} \tag{10.8}$$

Note that equation (10.5) that describes the evolution of a homogeneous scalar field is very similar to a harmonic oscillator.[27] The derivative of the scalar potential acts like a driving force and the expansion of our universe leads to the friction term $3H\dot{\phi}$.

10.2 Slow-roll inflation

As we discussed in the previous section, the successful models of inflation have a very flat potential along which the scalar field rolls down towards a minimum of the potential. If the scalar field rolls so slow, that we can neglect $\dot{\phi}^2$ compared to the potential value $V(\phi)$, then the scalar field behaves approximately like a cosmological constant, $\rho_\phi \approx -P_\phi$, and the universe undergoes a period of exponential expansion. This is called *slow-roll inflation* since the inflaton is rolling down the potential very slowly. Because matter and radiation will be diluted away due to this exponential expansion, we can neglect any other source of energy density and simply focus on the scalar field. We will get back to setting the initial conditions for our hot big bang scenario in subsection 11.3, when we discuss the end of inflation and the reheating of our universe.

There are two small (dimensionless) parameters that allow us to make the condition of a slowly rolling scalar field more precise. Recall that the first Friedmann equation

[27] If we choose $V(\phi) = \frac{1}{2}m^2\phi^2$, then the equation of motion is identical to a harmonic oscillator with friction.

sourced by a homogeneous scalar field takes the form

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = H^2 = \frac{8\pi G}{3}\rho_\phi = \frac{1}{3M_P^2}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right) . \tag{10.9}$$

So, we see that for $\dot{\phi}^2 \ll V(\phi)$ we have a nearly constant potential $V(\phi)$ value since the scalar field is only changing very slowly in time. This then implies a nearly constant Hubble parameter, so that it is very useful to introduce the dimensionless slow-roll parameter

$$\epsilon \equiv -\frac{\dot{H}}{H^2} . \tag{10.10}$$

Note that during a period of almost exponential expansion $\dot{H} < 0$ so that $\epsilon > 0$ (see equation (10.16) below). A period of inflation requires $\epsilon \ll 1$. Since we need inflation to last sufficiently long, we need ϵ not to change that quickly which is captured by the second dimensionless slow-roll parameter

$$\eta \equiv \frac{\dot{\epsilon}}{H\epsilon} . \tag{10.11}$$

This parameter keeps track of the relative change $\dot{\epsilon}/\epsilon$ per Hubble time and also needs to be small for an extended period of inflation.

Last time we have already used the idea of e-folds that measure the number of exponential expansions (to basis e) of our universe. We already defined the total number of e-folds as N_e but it is often more useful to measure time in terms of the number of e-folds. To this end we define

$$dN \equiv d\ln(a) = Hdt . \tag{10.12}$$

The total number of e-folds N_e is then given by

$$N_e = \int_{a_i}^{a_f} d\ln a = \ln\left(\frac{a_f}{a_i}\right) = \int_{t_i}^{t_f} Hdt \approx H_{inf}(t_f - t_i) , \tag{10.13}$$

where H_{inf} is the approximately constant Hubble parameter during inflation.

The second Friedmann equation

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi G}{3}(\rho(t) + 3P(t)) = -\frac{1}{6M_P^2}(\rho(t) + 3P(t)) , \tag{10.14}$$

in the presence of only a homogeneous scalar field gives

$$\dot{H} + H^2 = -\frac{1}{3M_P^2}(\dot{\phi}^2 - V(\phi)) . \tag{10.15}$$

Using equation (10.9) we then get

$$\dot{H} = -\frac{\dot{\phi}^2}{2M_P^2} . \tag{10.16}$$

Plugging this into the definition of ϵ we find

$$\epsilon = \frac{\dot{\phi}^2}{2M_P^2 H^2} . \tag{10.17}$$

Taking the time derivative gives

$$\dot{\epsilon} = \frac{\ddot{\phi}\dot{\phi}}{M_P^2 H^2} - \frac{\dot{\phi}^2 \dot{H}}{M_P^2 H^3}\,. \tag{10.18}$$

We can use this to rewrite η

$$\eta \equiv \frac{\dot{\epsilon}}{H\epsilon} = \left(\frac{\ddot{\phi}\dot{\phi}}{M_P^2 H^2} - \frac{\dot{\phi}^2 \dot{H}}{M_P^2 H^3}\right)\frac{2M_P^2 H}{\dot{\phi}^2} = 2\frac{\ddot{\phi}}{\dot{\phi}H} - 2\frac{\dot{H}}{H^2} = 2\frac{\ddot{\phi}}{\dot{\phi}H} + 2\epsilon\,. \tag{10.19}$$

10.2.1 The slow-roll equations

So far, we have not really made any approximations but since during inflation ϵ and η are very small we can calculate them and everything else to leading order to get rather simple expressions. For example, a small $\epsilon \ll 1$ (see eq. (10.17)) implies that $\dot{\phi}^2/(2M_P^2) \ll H^2$. It then follows from equation (10.9) that during slow-roll inflation

$$H^2 \approx \frac{V}{3M_P^2}\,. \tag{10.20}$$

This means that the Hubble constant during inflation is set by the value of the scalar potential. Since the scalar field is slowly rolling the value of the potential is only changing very slowly and therefore the Hubble constant is approximately constant during inflation. Similarly, a small $|\eta|$ and ϵ implies due to equation (10.19) that $\ddot{\phi} \ll \dot{\phi}H$. It then follows from equation (10.5) that

$$3H\dot{\phi} \approx -V'(\phi)\,. \tag{10.21}$$

Taking the time derivative of the above equation we get

$$3\dot{H}\dot{\phi} + 3H\ddot{\phi} \approx -\dot{\phi}V''(\phi)\,. \tag{10.22}$$

Using the equations (10.20), (10.21) and (10.22), we find the following approximate expression for ϵ and η

$$\begin{aligned} \epsilon &= \frac{\dot{\phi}^2}{2M_P^2 H^2} \approx \frac{(V')^2}{18 M_P^2 H^4} = \frac{M_P^2}{2}\left(\frac{V'}{V}\right)^2, \\ \eta &= 2\frac{\ddot{\phi}}{\dot{\phi}H} + 2\epsilon \approx \frac{2}{\dot{\phi}H}\left(-\frac{\dot{\phi}V''}{3H} - \frac{\dot{H}\dot{\phi}}{H}\right) + 2\epsilon = -2M_P^2\frac{V''}{V} + 2M_P^2\left(\frac{V'}{V}\right)^2 \end{aligned} \tag{10.23}$$

So, we see that we can express ϵ and η entirely in terms of the scalar potential. It is convenient to introduce the slow-roll parameters ϵ_V and η_V that are defined by

$$\epsilon_V \equiv \frac{M_P^2}{2}\left(\frac{V'(\phi)}{V(\phi)}\right)^2 \approx \epsilon\,, \tag{10.24}$$

$$\eta_V \equiv M_P^2 \frac{V''(\phi)}{V(\phi)} \approx -\frac{1}{2}\eta + 2\epsilon\,. \tag{10.25}$$

The smallness of ϵ_V is equivalent to the condition that the first derivative of the potential is small compared to the value of the potential, while the smallness of $|\eta_V|$ is equivalent to the smallness of the second derivative of the potential. It follows from the equations (10.23), that the slow-roll conditions $\epsilon, |\eta| \ll 1$ are equivalent to $\epsilon_V, |\eta_V| \ll 1$.

We can also express the number of e-folds of inflation in terms of the slow-roll parameter ϵ_V. Using equation (10.17) we can rewrite

$$Hdt = \frac{H}{\dot{\phi}} d\phi = \frac{1}{\sqrt{2\epsilon}} \frac{|d\phi|}{M_P} \approx \frac{1}{\sqrt{2\epsilon_V}} \frac{|d\phi|}{M_P} \,. \tag{10.26}$$

Now we use this in the definition of the number of e-folds given in equation (10.13) to get

$$N_e = \int_{t_i}^{t_f} Hdt \approx \int_{\phi_i}^{\phi_f} \frac{1}{\sqrt{2\epsilon_V}} \frac{|d\phi|}{M_P} = \frac{1}{M_P^2} \left| \int_{\phi_i}^{\phi_f} \frac{V(\phi)}{V'(\phi)} d\phi \right| \,. \tag{10.27}$$

10.3 Examples of inflationary models

10.3.1 Natural inflation

After discussing all the relevant equations, let us now discuss a concrete model of inflation that is called natural inflation. In this model the inflaton field has a discrete shift symmetry $\phi \to \phi + 2\pi f$. The potential that respects this shift symmetry is given by

$$V(\phi) = \lambda^4 \left[1 + \cos\left(\frac{\phi}{f}\right) \right] \,. \tag{10.28}$$

Here we have set the minimum value of the potential equal to zero since the current cosmological constant is so small that it would not matter for the period of inflation, if we add to this potential a constant that is $10^{-120} M_P^4$ or not. One period of the potential is shown in figure (44).

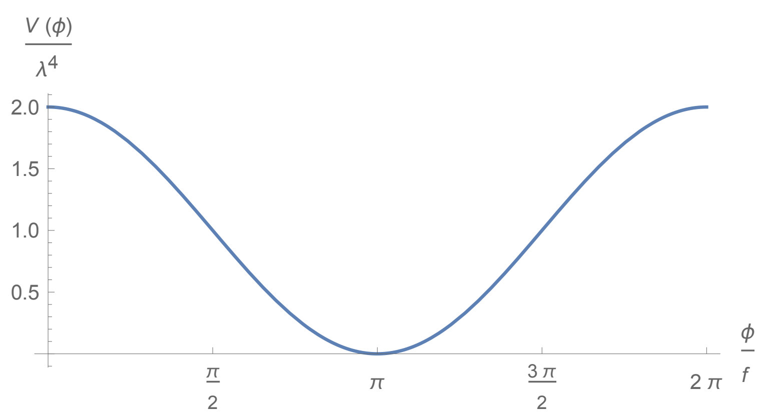


Figure 44: The scalar potential for natural inflation.

Now we calculate the slow-roll parameters

$$\epsilon_V = \frac{M_P^2}{2f^2}\left(\frac{\sin\left(\frac{\phi}{f}\right)}{1+\cos\left(\frac{\phi}{f}\right)}\right)^2 , \tag{10.29}$$

$$\eta_V = -\frac{M_P^2}{f^2}\frac{\cos\left(\frac{\phi}{f}\right)}{1+\cos\left(\frac{\phi}{f}\right)} . \tag{10.30}$$

Since the scalar potential is zero (or very small) at the minimum of the potential, while the second derivative is large, we see that we cannot satisfy $|\eta_V| \ll 1$ near the minimum of the potential.[28] Away from the minimum at $\phi = \pi f$, there is no enhancement of ϵ_V and η_V from the denominator but we don't have any substantial suppression from the nominator either, since it is not possible to have $|\sin(\phi/f)| \ll 1$ and at the same time $|\cos(\phi/f)| \ll 1$. So, the only way to get small ϵ_V and $|\eta_V|$ in this model is by choosing $f \gg M_P$. This automatically suppresses both slow-roll parameters and allows for a period of slow-roll inflation.

We can calculate the number of e-folds in this model using equation (10.27)

$$N_e \approx \int_{\phi_i}^{\phi_f} \frac{1}{\sqrt{2\epsilon_V}}\frac{|d\phi|}{M_P} = \frac{f}{M_P^2}\int_{\phi_i}^{\phi_f}\left|\frac{1+\cos\left(\frac{\phi}{f}\right)}{\sin\left(\frac{\phi}{f}\right)}\right||d\phi| = 2\frac{f^2}{M_P^2}\left|\log\left(\frac{\sin\left(\frac{\phi_f}{2f}\right)}{\sin\left(\frac{\phi_i}{2f}\right)}\right)\right| . \tag{10.31}$$

We see, probably as one would expect, that the number of e-folds diverges as one moves the starting point ϕ_i closer and closer to a maximum $\phi \in 2\pi n f$, $n \in \mathbb{Z}$.

The end of inflation is defined as the point at which one of the slow-roll parameters becomes equal to 1. Usually, the expansion after that point can only add some order one number of e-folds. So, this end of inflation is not the exact end of the exponential expansion but a rough guide. In our example ϵ_V and $|\eta_V|$ are of the same order and which one is bigger depends on how close to the minimum we are. For simplicity we just focus on $\epsilon_V = 1$. One finds that

$$\epsilon_V(\phi_f) = 1 \qquad \Leftrightarrow \qquad \phi_f = f\left[\pi \pm \arctan\left(\frac{2\sqrt{c}}{c-1}\right)\right] , \tag{10.32}$$

where $c = \sqrt{2}f/M_P$. Since a period of slow-roll requires $f \gg M_P$, we have that $c \gg 1$ and we can approximate

$$\arctan\left(\frac{2\sqrt{c}}{c-1}\right) \approx \frac{2}{\sqrt{c}} = 2^{\frac{3}{4}}\sqrt{\frac{M_P}{f}} . \tag{10.33}$$

Now if we take for example $f = 100 M_P \gg M_P$, then we have $\phi_f \approx 2.97 f = 297 M_P$ (or $\phi_f \approx 2\pi f - 2.97 f \approx 3.31 f = 331 M_P$) and we find for the number of e-folds

$$N_e \approx 2\times 10^4\left|\log\left(\frac{\sin\left(\frac{\pi}{2}\pm\frac{2^{\frac{3}{4}}}{20}\right)}{\sin\left(\frac{\phi_i}{200 M_P}\right)}\right)\right| \approx -2\times 10^4 \log\left|\sin\left(\frac{\phi_i}{200 M_P}\right)\right| - 71 . \tag{10.34}$$

[28]The second derivative of the scalar potential at the minimum determines the mass of the inflaton today. Since we don't observe any very light scalar fields, the second derivative at the minimum has to be much larger than the value of the potential.

If we want to get 60 e-folds, then we can numerically solve for ϕ_i and find $\phi_i \approx 2.91f = 291M_P$ (or $\phi_i \approx 2\pi f - 2.91f \approx 337M_P$). We see that in this example ϕ travels a distance in field space that is larger than M_P. Models of this type in which $\Delta\phi \equiv |\phi_i - \phi_f| \gtrsim M_P$ are called *large field models*. Such large field models are currently being tested by observations and are already highly constrained. Constructing these models in a controlled way provides many theoretical challenges (and ideally requires a full-fledged theory of quantum gravity). In these large field models ϵ_V is much larger than in small field models. This in turn implies that large field models require less tuning of the scalar potential in order to get a period of inflation.

10.3.2 $m^2\phi^2$ inflation

As we also see from above, since $\phi_i \approx 2.91f$ and $\phi_f \approx 2.97f$, the inflaton field does not really explore much of the potential but stays very close to the minimum. In such cases we can just expand the potential around the minimum and find

$$V(\phi) = \lambda^4 \left[1 + \cos\left(\frac{\phi}{f}\right)\right] \approx \frac{\lambda^4}{2f^2}\left((\phi - \pi f)^2 + \mathcal{O}\left((\phi - \pi f)^4\right)\right) . \tag{10.35}$$

Keeping only the leading term, defining $m^2 = \lambda^4/f^2$ and shifting $\phi \to \phi + \pi f$ we find the potential

$$V(\phi) = \frac{1}{2}m^2\phi^2 . \tag{10.36}$$

This is arguably the simplest potential for large field inflation and it is very generic in the sense that for every massive scalar field the potential around the minimum is quadratic (since $m^2 \propto V''(\phi)$).

In this simple model the slow-roll parameters are

$$\epsilon_V = \eta_V = 2\left(\frac{M_P}{\phi}\right)^2 . \tag{10.37}$$

So, as long as $\phi \gg M_P$ we have slow-roll inflation and inflation ends when $\phi \approx \sqrt{2}M_P$. The number of e-folds in this model is given by

$$N_e \approx \int_{\phi_i}^{\phi_f} \frac{1}{\sqrt{2\epsilon_V}}\frac{|d\phi|}{M_P} = \left|\int_{\phi_i}^{\sqrt{2}M_P} \frac{\phi d\phi}{2M_P^2}\right| = \frac{\phi_i^2}{4M_P^2} - \frac{1}{2} . \tag{10.38}$$

For sixty e-folds we need

$$\phi_i = 2M_P\sqrt{60.5} \approx 15.6M_P . \tag{10.39}$$

Plugging this into ϵ_V we find

$$\epsilon_V(\phi_i) = 2\left(\frac{M_P}{\phi_i}\right)^2 = \frac{1}{121} \approx .0083 . \tag{10.40}$$

The bound from the Planck satellite from February 2015 excluded this model at the 2σ confidence level by providing the upper bound $\epsilon_V \leq .0069$.

The above more general model of natural inflation has essentially also been excluded at the 2σ level by the Planck 2018 data release that is based on a refined analysis and the combination with other experiments like BICEP and the Keck array. This shows that this a very active research field with constantly improving observational constraints.

10.3.3 Small field models of inflation

By definition small field models of inflation satisfy $\Delta\phi = |\phi_i - \phi_f| \ll M_P$. As an example, let us again return to the model above and now focus on the region near a maximum, for example near $\phi = 0$. In this case we can expand all the equations around $\phi = 0$ and keep the leading terms. This gives us an example of so-called hilltop inflation, in which the period of inflation happens near an unstable point of the potential.

The potential becomes

$$V(\phi) = \lambda^4 \left[2 - \frac{1}{2}\left(\frac{\phi}{f}\right)^2\right], \tag{10.41}$$

which is unbounded from below but we of course only consider it as an approximation near the ϕ values for which inflation takes place. This potential is then only valid for $\phi/f \ll 1$ and does not need to be completed to a full cosine but can rather have arbitrary higher order corrections of $\mathcal{O}\left((\phi/f)^3\right)$.

The slow-roll parameters are

$$\epsilon_V = \frac{M_P^2}{2}\left(\frac{\frac{\phi}{f^2}}{2}\right)^2 = \frac{M_P^2\phi^2}{8f^4}, \tag{10.42}$$

$$\eta_V = -\frac{M_P^2}{2f^2}. \tag{10.43}$$

We see that both are small as long as $M_P, \phi \ll f$. We assume that corrections to the potential will modify ϵ_V and η_V near $\phi \sim f/100$ and that inflation ends around this point. The number of e-folds is then given by

$$N_e \approx \int_{\phi_i}^{\phi_f} \frac{1}{\sqrt{2\epsilon_V}}\frac{|d\phi|}{M_P} = \frac{2f^2}{M_P^2}\int_{\phi_i}^{f/100}\frac{d\phi}{\phi} = \frac{2f^2}{M_P^2}\log\left(\frac{f}{100\,\phi_i}\right). \tag{10.44}$$

For example, for concreteness we can choose $f = 5M_P$ so that 60 e-folds require

$$N_e = 60 \approx 50\log\left(\frac{M_P}{20\,\phi_i}\right) \qquad \Leftrightarrow \qquad \phi_i \approx .015 M_P \ll M_P. \tag{10.45}$$

This leads to

$$\epsilon_V(\phi_i) \approx 5\times 10^{-8} \qquad \Leftrightarrow \qquad M_P|V'(\phi_i)| \approx 2\times 10^{-4}\, V(\phi_i), \tag{10.46}$$

which is a larger fine tuning of the potential as one would require in large field models.

Summary: Inflation

In this section we have learned about single field, slow-roll inflation that is currently the most promising candidate of inflation. In particular, we have seen that slow-roll inflation is tied via the slow-roll parameters

$$\epsilon_V \equiv \frac{M_P^2}{2}\left(\frac{V'(\phi)}{V(\phi)}\right)^2 , \tag{10.47}$$

$$\eta_V \equiv M_P^2 \frac{V''(\phi)}{V(\phi)} , \tag{10.48}$$

to the scalar potential $V(\phi)$. Slow-roll inflation takes place whenever $\epsilon_V, \eta_V \ll 1$.
We have seen that there exist so-called large field and small field inflationary models. The large field inflationary models are currently being tested by experiments and the probably simplest model, $V(\phi) = \frac{1}{2}m^2\phi^2$, has been excluded only a few years ago. However, there are many more possible shapes for the inflationary potential and currently experiments are restricting the possible scalar potentials further and further, as we will discuss more in the next section.

11 Experimental constraints and reheating

In the previous section we discussed the relevant equations for slow-roll inflation as well as a few concrete slow-roll models. In this section we discuss a little bit the experimental status of these models, the reheating process and for the first time in this course the deviation from a completely homogeneous universe.

11.1 Experimental constraints on inflationary models

As we mentioned in the previous section and as you have seen in the homework, experiments place bounds on existing inflationary models. Several models have already been excluded and future experiments will tighten the bounds and further shrink the parameter space for inflationary models. Ideally this will ultimately single out one particular model of inflation.

There are several ground-based experiments that observe a small patch on the sky and have provided lots of important data that constraints the cosmological parameters. Satellites on the other hand have access to most of the sky, however they require that the entire measuring apparatus can be transported into space where it has to work without being maintained or upgraded. We focus here on the bounds on inflationary models that were released by the Planck collaboration in February 2015 and the BICEP2 and Keck array in October 2015. The Planck satellite has measured the black body spectrum from the CMB at several different frequencies and also the polarization of the photons. Its data tightens most bounds and favors simple single field slow-roll inflationary models. Furthermore, by constraining the slow-roll parameters

$$\epsilon_V \equiv \frac{M_P^2}{2}\left(\frac{V'(\phi)}{V(\phi)}\right)^2 , \tag{11.1}$$

$$\eta_V \equiv M_P^2 \frac{V''(\phi)}{V(\phi)} , \tag{11.2}$$

it tells us about the form of the potential at a time of 50-60 e-folds before the end of inflation.

The experimental data is usually presented in terms of constrains on the so-called spectral index $n_s = 1 - 6\epsilon_V + 2\eta_V$ and the tensor-to-scalar ratio $r = 16\epsilon_V$. While inflationary models generically predict values of $n_s \neq 1$, these values can in principle be larger or smaller than 1. However, the data clearly requires $n_s < 1$. The particular case of $\epsilon_V = 0$, $\eta_V > 0$ which gives rise to $n_s > 1$ is experimentally excluded. This case would have corresponded to a true de Sitter phase, i.e., a positive cosmological constant, instead of a slowly rolling scalar field. The figure below shows the experimental constraints in the (n_s, r)-plane together with a variety of inflationary models:

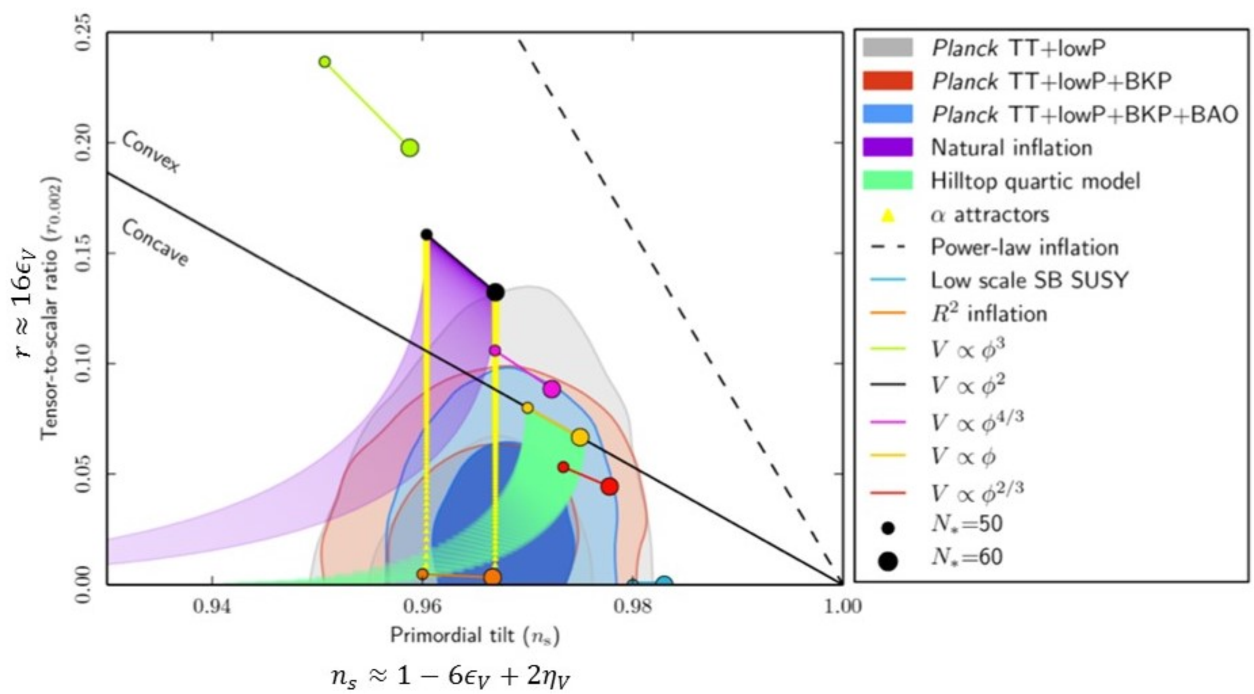


Figure 45: Marginalized joint 68% and 95% CL regions for n_s and $r_{0.002}$ from *Planck* 2015 in combination with other data sets, compared to theoretical predictions of selected inflationary models.

The most stringent bounds are the dark blue region (68% confidence level (CL)) and the light blue region (95% CL). Usually, a model is not considered to be excluded, if it is within the 2σ (i.e., 95% CL) region. Once it is outside this region it is somewhat or very unlikely depending on how far outside it is.

Theoretical models are shown in this plot as dots connected by a line or as a band. The two dots that are shown for particular models, like for example the two black dots for $V \propto \phi^2$ correspond to $N_e = 50$ and $N_e = 60$ and the line connecting the dots correspond to N_e values between 50 and 60. The reason that this range is shown is that we do not know for sure the exact number of e-folds N_e that we can observe.

In the sense discussed above, the simplest model of inflation, the $m^2\phi^2$ model we discussed in subsection 10.3.2, became very unlikely in 2015 and other models seem favored by the data. For example, as you have seen in the homework, models with $V \propto \phi^p$ are more favored for $p < 2$. For the model of natural inflation that we discussed in subsection 10.3.1 we see a purple band. The reason for this is that this model has the free parameter f and depending on the value of this parameter we get different predictions. So, f in this model is constraint by the data to take certain values. We also see that the purple band is connected to the $m^2\phi^2$ model that we discussed previously as a limiting case of natural inflation.

Note that the parameter r determines directly the energy scale during inflation via the equation $V_{inf} \approx 2 \times 10^{16} GeV \left(\frac{r}{.1}\right)^{\frac{1}{4}}$. While experiments so far have only placed an upper bound $r < .07$ (95% CL), any future detection of a non-zero r would tell us directly the energy scale at which inflation took place.

The figure above shows a few more models of inflation that are just a very small selection of the actual models that have been proposed. Hopefully in the future the ever-improving observations will narrow the parameter space down to a single model.

11.2 Beyond slow-roll single field inflation

As you can already guess from the above plot, there are a lot of single field slow-roll models since we can tune the parameters in many scalar potentials such that they contain sufficiently flat regions. In addition to these single field slow-roll models there are also multifield models in which we don't just have a single scalar field but many. These models are obviously much more complicated and make a variety of interesting predictions. None of these have been observed so far so that multifield models are constrained (but by no means excluded).

In addition to these slow-roll models in which a scalar field is rolling very slowly, it is also possible to get a period of accelerated expansion from a very fast rolling scalar field. This seemingly counterintuitive statement follows from the fact that we can't neglect higher order corrections to the scalar field's kinetic term, once the field is fast rolling, i.e., terms like $(\partial_\mu\phi\partial^\mu\phi)^n$ for $n > 1$ become important. If these higher order corrections take a specific form, then the fast-rolling scalar field leads to a period of inflation. This class of models is usually called K-inflation, where K stands for 'kinetic'.

11.3 Reheating

Inflation solves several problems in our universe but it also drastically modifies the evolution of our early universe. In particular, the rapid expansion during inflation leads to an essentially empty universe since the contributions to the energy density from matter, radiation and curvature scale with a negative power of the scale factor $a(t)$. Since $a(t)$ grows by a huge factor, essentially the entire energy density of the universe is given by the contribution of the scalar field:

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi)\,, \tag{11.3}$$

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi)\,. \tag{11.4}$$

At the end of inflation one of the slow-roll parameters becomes of order 1, which is equivalent to the kinetic term $\dot{\phi}^2$ becoming important. If the scalar field would oscillate around the minimum at which $V(\phi) \approx 0$, then we would have $\rho_\phi \approx P_\phi$, i.e., a phase with $w = 1$. In that case it follows from equation (2.1) that $\rho_\phi \propto a(t)^{-3(1+w)} = a(t)^{-6}$. The universe at this point keeps expanding since the first Friedmann equation becomes

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{\dot{\phi}^2}{6M_P^2} > 0\,. \tag{11.5}$$

This would mean that the energy density of the universe would go to zero and we would be left with a completely empty universe. Likewise, if the inflaton oscillates around the minimum of the potential and the Hubble friction is negligible (because H becomes small at the end of inflation), then the kinetic energy and the potential energy are on average the same $\frac{1}{2}\langle\dot{\phi}^2\rangle = \langle V(\phi)\rangle$ and the inflaton behaves like pressure-less matter $\langle P_\phi\rangle = 0$. In this case we have $\rho_\phi \propto a(t)^{-3}$. So, what we need is a way to transfer the kinetic energy of the scalar field into the standard model degrees of freedom so that our hot big bang scenario can take place.

The period during which this energy transfer is happening is called reheating, since the universe during inflation has essentially zero temperature and then the universe gets reheated.

Worked problem 11.1: Temperature decrease during inflation

Assume our universe undergoes a phase of inflation which increases its size by 60 e-folds and that starts at $T = 10^{16}GeV$. What is the temperature at the end of inflation?
Solution: We know that $T \propto 1/a(t)$. So, if $a(t)$ grows by $e^{60} \approx 10^{23}$ then the temperature reduces to $T \approx 10^{-7}GeV = 10^2 eV$. This is much smaller than the temperature we need to establish thermal equilibrium due to ample interaction, i.e., $T \gg 100GeV$. It is also smaller than the temperature of $1MeV$ around which nucleosynthesis happens. We therefore expect the reheating that is described in this section to have taken place. Compared to the reheating temperature $.01GeV < T_{reh} < 10^{16}GeV$ the above temperature is essentially zero.

The reheating of the universe can be accomplished by coupling the inflaton scalar field ϕ to other fields like for example the fermions ψ in the standard model via $\phi\bar{\psi}\psi$ Yukawa couplings in the action. These then lead to an extra term in the equation of motion of the inflaton

$$\ddot{\phi} + 3H\dot{\phi} + \Gamma_\phi\dot{\phi} + V'(\phi) = 0\,. \tag{11.6}$$

The additional friction term Γ_ϕ is proportional to the decay width of the inflaton that in turn is determined by the prefactor of the above Yukawa coupling. During inflation we need $H \gg \Gamma_\phi$ so that the above term does not spoil inflation. At the end of inflation H decreases and then the term Γ_ϕ will become important and the energy transfer from the inflaton to the standard model particles will take place.

This leads then to a reheating of the universe. Note that the spatial inhomogeneities of the inflaton field $\nabla^2\phi/a(t)^2$ have been sufficiently suppressed during inflation to avoid the horizon problem and the reheating will naturally lead to a homogeneous temperature distribution. If the inflaton only couples to some of the standard model particles, then the ample interactions that take place will lead to the required initial conditions in which all standard model particles are in thermal equilibrium.

During inflation the scalar field has essentially only potential energy that then gets transferred into kinetic energy towards the end of inflation. This kinetic energy then gets in turn transferred into the standard model particles as soon as the Hubble parameter has decreased sufficiently at the end of inflation so that $H \lesssim \Gamma_\phi$. This means that the maximum reheating temperature is set by the energy scale during inflation $T_{reh,max} \sim V_{inf}^{\frac{1}{4}}$. This energy scale during inflation is currently constrained by observation to be smaller than roughly the GUT scale, which is $10^{16} GeV$. This means in turn that the reheating temperature cannot be larger than the GUT scale so that after inflation there is no problem with relics that might appear during the breaking of the GUT gauge group to the standard model gauge groups.

The lower end for the reheating temperature and therefore also for the energy density during inflation is essentially only constrained by our observation of the correct abundance of light elements like Helium and Hydrogen. As we discussed, this can be correctly explained by the nucleosynthesis which starts around a few MeV. So, although we in principle understand physics up to much larger energy scales of $1 TeV$, it is possible that at the end of inflation the universe only gets reheated to a temperature of a few MeV and then the thermal evolution we discussed takes place. This means that there is a huge energy range $.001\, GeV \lesssim V_{inf}^{\frac{1}{4}} \lesssim 10^{16}\, GeV$ during which inflation could have happened.

Note that the reheating temperature T_{reh} does not have to be equal to the energy density at the end of inflation but is only bounded by it from above. It is possible that the reheating temperature is lower than the energy density during inflation. However, the exact relation requires a specific model of inflation with a specific reheating mechanism.

11.4 The inhomogeneous universe

So far in these lecture notes we have restricted ourselves to a completely homogeneous and isotropic universe. While this is a very good approximation for our universe on large scales, it is certainly not true on smaller scales like the size of galaxies. How can we explain these observed inhomogeneities and thereby our very own existence?

A natural thought might be that the unknown initial conditions of our universe were not perfectly homogeneous and through time gravitational clumping leads to structure formation. A problem with this is that a sufficiently long period of inflation erases any initial inhomogeneities. So, it seems that a period of inflation seems to make the existence of inhomogeneities and structure formation more difficult. However, this is not the case. The reason is Heisenberg's uncertainty principle that applies to quantum mechanics and also to quantum field theory. Since our universe is governed by a quantum theory, we cannot just treat the inflaton field as a classical scalar field but we have to take quantum mechanics into account. Heisenberg's uncertainty principle says that we cannot exactly determine the value of the field ϕ and its conjugate momentum. This means that we should allow for small fluctuations in ϕ:

$$\phi = \bar{\phi}(t) + \delta\phi(t, \vec{x})\,. \tag{11.7}$$

Here we have decomposed the field ϕ into a background field $\bar{\phi}(t)$ that is homogeneous and satisfies the classical equations of motion and a small fluctuation that depends on space and time, i.e., that is not homogeneous. Likewise, we can perturb the metric $g_{\mu\nu}$ and the energy density and pressure. As long as these perturbations are small, we can derive their equations of motion by linearizing the equations of motion. We can then study these linear equations of motion and quantize the fluctuations. We can track the evolution of these fluctuation and derive their features. While this is not too complicated it is somewhat technical and a complete treatment would take a few more sections so we will skip the details and focus on the results.

The quantum fluctuations of the inflaton field mean that the position of the inflaton field is different at different points in space (since $\delta\phi$ depends on $\vec{x}$). This is sketched in figure 46.

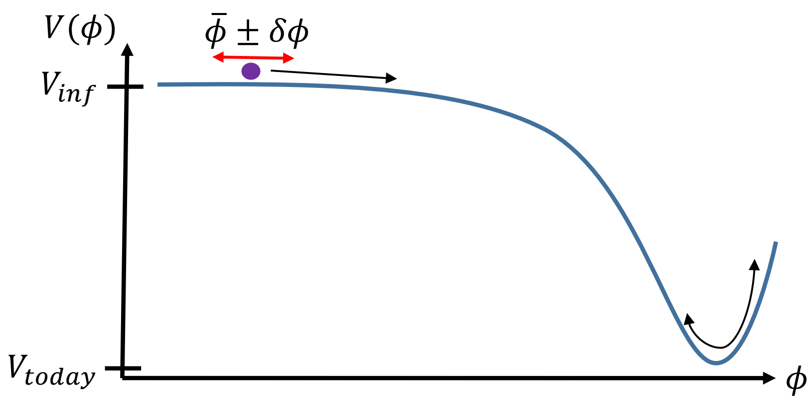


Figure 46: The position ϕ of the scalar field changes in space and time due to quantum fluctuations $\delta\phi(t, \vec{x})$.

What this means, and what one finds by solving the equations for the perturbations, is that inflation ends at slightly different times at different points in space. This then leads to slightly more or less dense regions at the end of inflation. We can explicitly calculate the evolution of the fluctuations during and after inflation as long as they are small and perturbation theory is applicable. What one finds is that small quantum fluctuations are getting stretched and amplified during inflation. At the end of inflation, the perturbations grow slowly during the radiation dominated epoch and then more rapidly during the matter dominated epoch. The gravitational attraction will eventually make the perturbations so large that structures like stars and galaxies will form. At this point the linear perturbation theory has broken down. However, the density fluctuations lead to small temperature fluctuations in the cosmic microwave background that we can observe and theoretically calculate using linearized perturbation theory. The latest experimental observation of these temperature fluctuations from the Planck satellite is shown in figure 47.

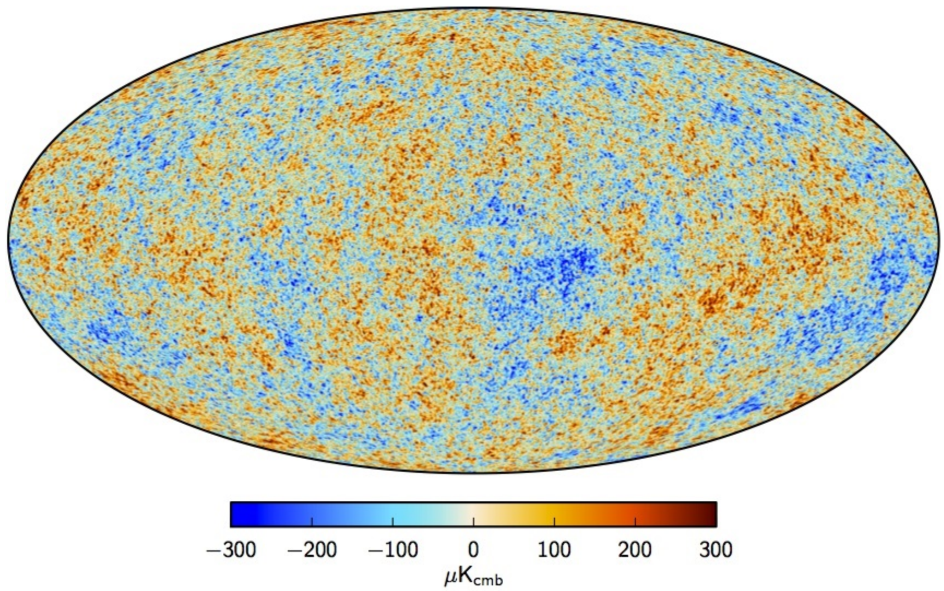


Figure 47: Very small temperature fluctuations in the CMB as observed by the Planck satellite.

The CMB photons that we measure come from a huge sphere that surrounds us and similar to a map of the surface of the earth we can project this sphere on a 2-dimensional surface, which is what is shown in the figure 47. The temperature $T = 2.725K$ is very homogeneous and the root mean square variations are only $18\mu K$ so that the variations in the temperature are roughly $\delta T/T \sim 10^{-5}$. These small fluctuations look like random noise and that is exactly what they are: random quantum noise. This noise nevertheless contains a tremendous amount of information. For initial quantum noise that evolves during the radiation and matter dominated era until the release of the CMB 380,000 years after the big bang, one can derive very precisely the expected 2-point function for the temperature fluctuations. This 2-point function depends only on the angle φ between the two points on the sky and we can expand it therefore in terms of Legendre Polynomials $P_l(\cos(\varphi))$. The resulting theoretical curve together with the data is shown in figure 48. We see that the data and theory beautifully agree with each other and that the random quantum noise indeed contains a structure that is not immediately visible in figure 47.

From the measured fluctuations we can determine the ratio of the inflaton potential $V(\phi)$ and the slow-roll parameter ϵ_ϕ 50 to 60 e-folds before the end of inflation. Concretely from the detailed perturbation theory calculation one finds for slow-roll models that

$$\frac{1}{24\pi^2}\frac{V(\phi)}{M_P^4}\frac{1}{\epsilon_V} \approx 2.2 \times 10^{-10} \,. \tag{11.8}$$

This equation allows us to fix a parameter in our inflationary model. For example, for $V(\phi) = \frac{1}{2}m^2\phi^2$ we calculated in subsection 10.3.2 for $N_e = 60$ that $\phi_i = 15.6M_P$ and

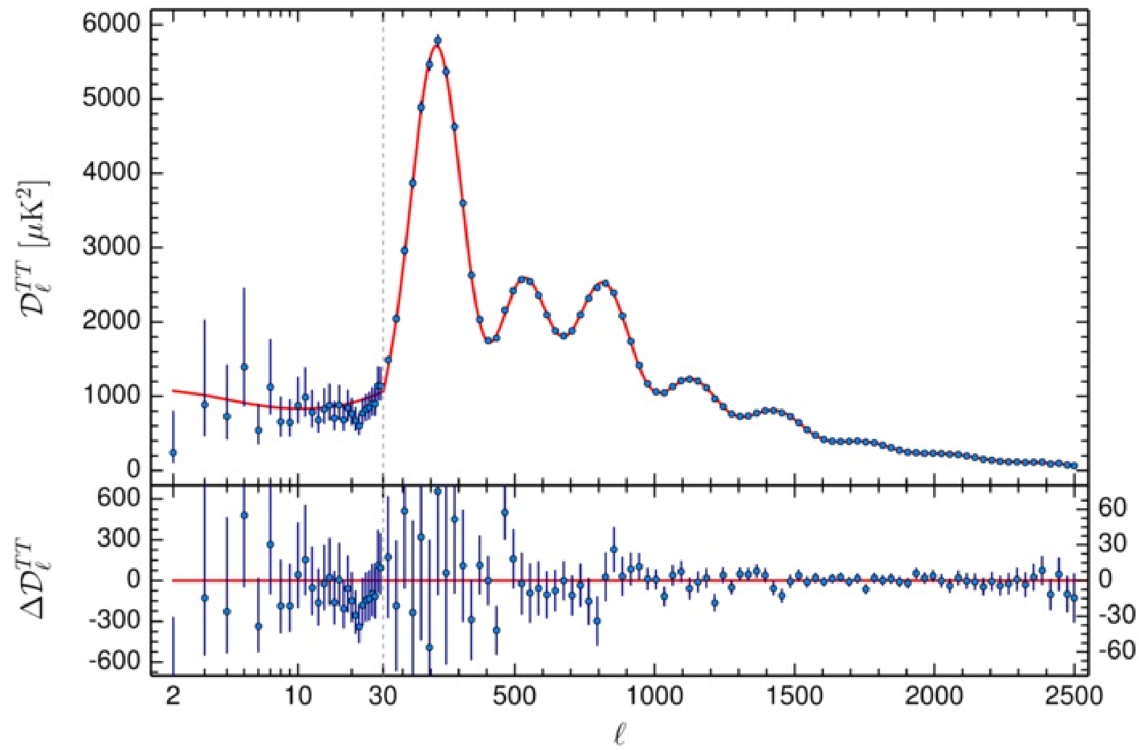


Figure 48: The 2-point function of the temperature fluctuations as a function of the multipole l. The lower part is zoomed in and shows the tiny differences between the red theory curve and the data with 1σ error bars.

$\epsilon_V = \frac{1}{121}$, so that

$$\frac{1}{24\pi^2}\frac{V(\phi)}{M_P^4}\frac{1}{\epsilon_V} = \frac{1}{48\pi^2}\frac{m^2(15.6M_P)^2}{M_P^4}121 \approx 62\frac{m^2}{M_P^2} \approx 2.2\times 10^{-10}\,. \tag{11.9}$$

Recalling that $M_P \approx 2.4\times 10^{18}GeV$ we find

$$m \approx 2.4\sqrt{\frac{2.2}{62}}10^{13}GeV \approx 4.5\times 10^{12}GeV\,. \tag{11.10}$$

So, we have found that the inflaton in this model is much heavier than the all the particles in the standard model that have masses up to only $173\,GeV$.

In addition to the temperature fluctuations in the CMB one can in principle also measure the polarization of the CMB photons. This would provide us with additional information about the perturbations that were generated during inflation. In particular the fluctuations in the space-time metric that are gravitational waves and that are generated during inflation give rise to particular patterns in the polarization of the CMB photons, as is shown in figure 49.

This is particularly interesting because contrary to the density/temperature fluctuations that are enhanced by a factor $1/\epsilon_V$ (see equation (11.8)), the so called tensor perturbations are not enhanced. Thus, they are much smaller and harder to measure but at the same time any detection would tell us the actual energy scale during inflation

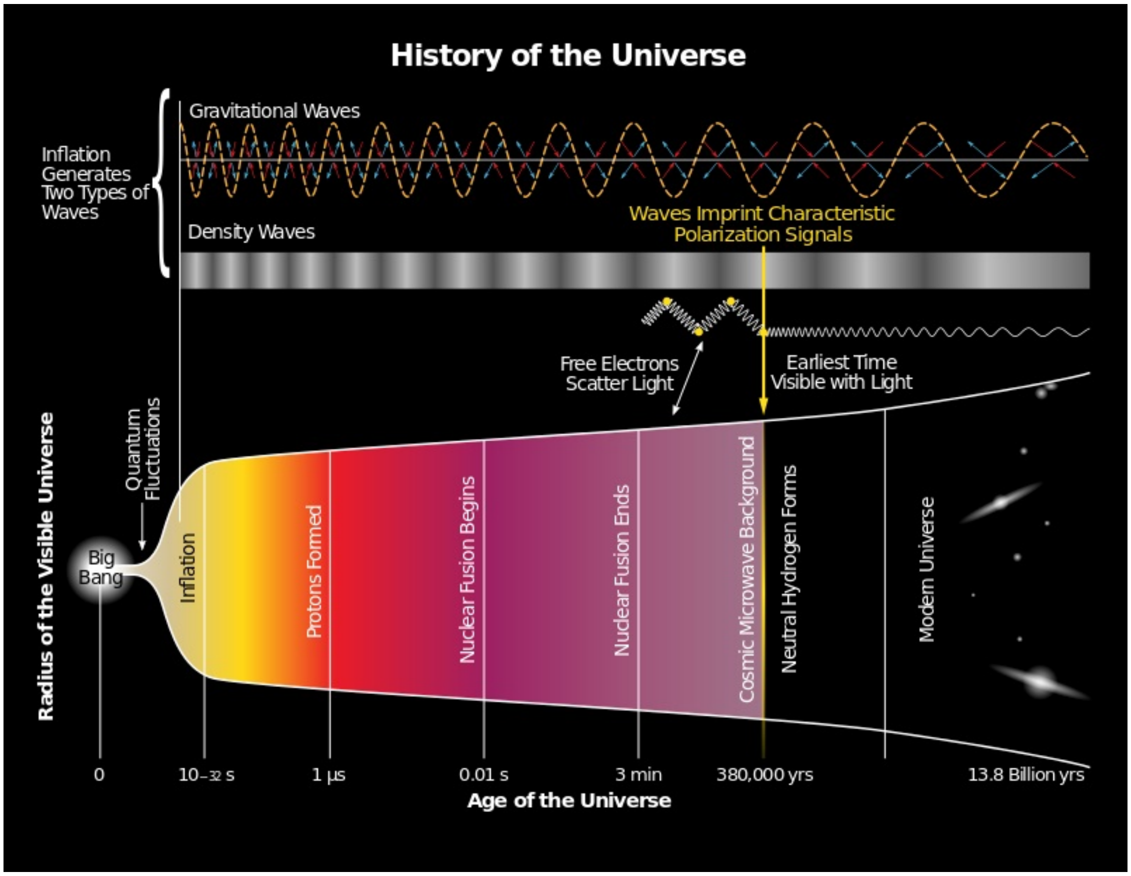


Figure 49: During inflation density and metric perturbations are generated. They can indirectly be measured by observing the temperature fluctuations and the polarization of the CMB photons.

(since the unknown enhancement from $1/\epsilon_V$ is absent). However, so far, we have not been able to detect this particular type of polarization in the CMB photons so that we only have an upper bound on the so-called scalar-to-tensor ratio $r < .07$ (95% CL). The energy scale during inflation is then determine through this parameter r as

$$V_{inf}^{\frac{1}{4}} \approx 2 \left(\frac{r}{.1}\right)^{\frac{1}{4}} 10^{16} GeV \,. \tag{11.11}$$

Future experiments can detect r or lower the bound potentially up to $r \gtrsim 10^{-4}$ so that any future detection of r would corresponds to a very high scale of inflaton. In particular this scale would be roughly 12 orders of magnitude larger than what can be tested in particle accelerators! This energy scale would also be very close to the Planck scale which might allow us to get insights into the theory of quantum gravity that governs our universe. So, the future of theoretical and observational cosmology holds the promise of great discoveries!

Summary: Reheating and quantum fluctuations

We started out by looking at some of the experimental bounds that are constraining large field models of slow-role inflation. There are ongoing and future experiments that constantly restrict the parameters further and further and hopefully will eventually zoom in on a particular model. We also discussed the requirement for reheating the universe at the end of inflation. The reheating process transfers kinetic energy from the inflaton particle to the particles in the standard model of particle physics. Since inflation has smoothed out any inhomogeneities the reheated universe is homogeneous and isotropic and from there our usual hot big bang description can take off.

Lastly, we discussed that quantum fluctuations of the inflaton field leads to inhomogeneities in the cosmic microwave background that we have observed after inflation was invented. This means that quantum fluctuations are the seeds for the first stars and galaxies in our universe!

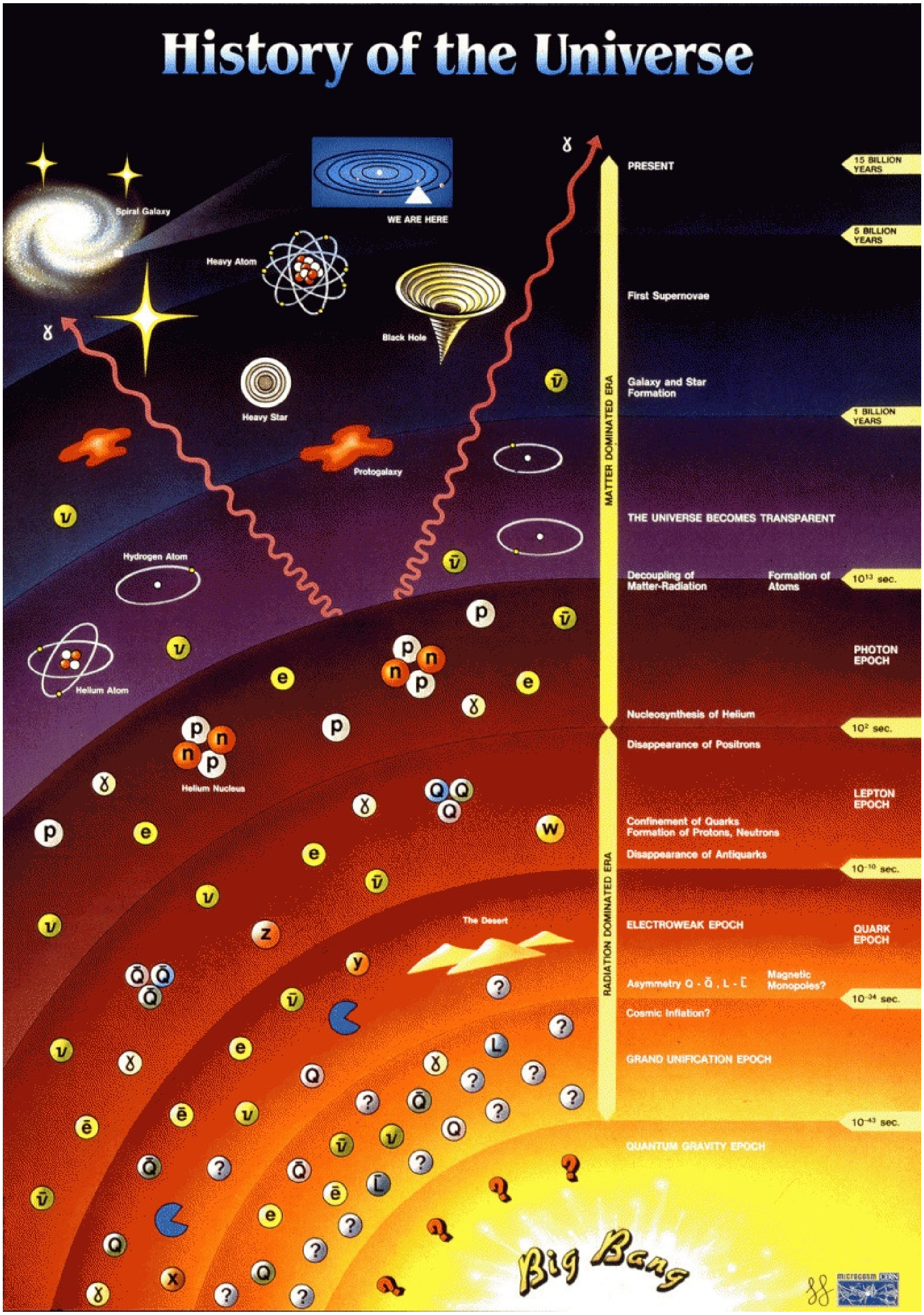
12 Summary: Our universe

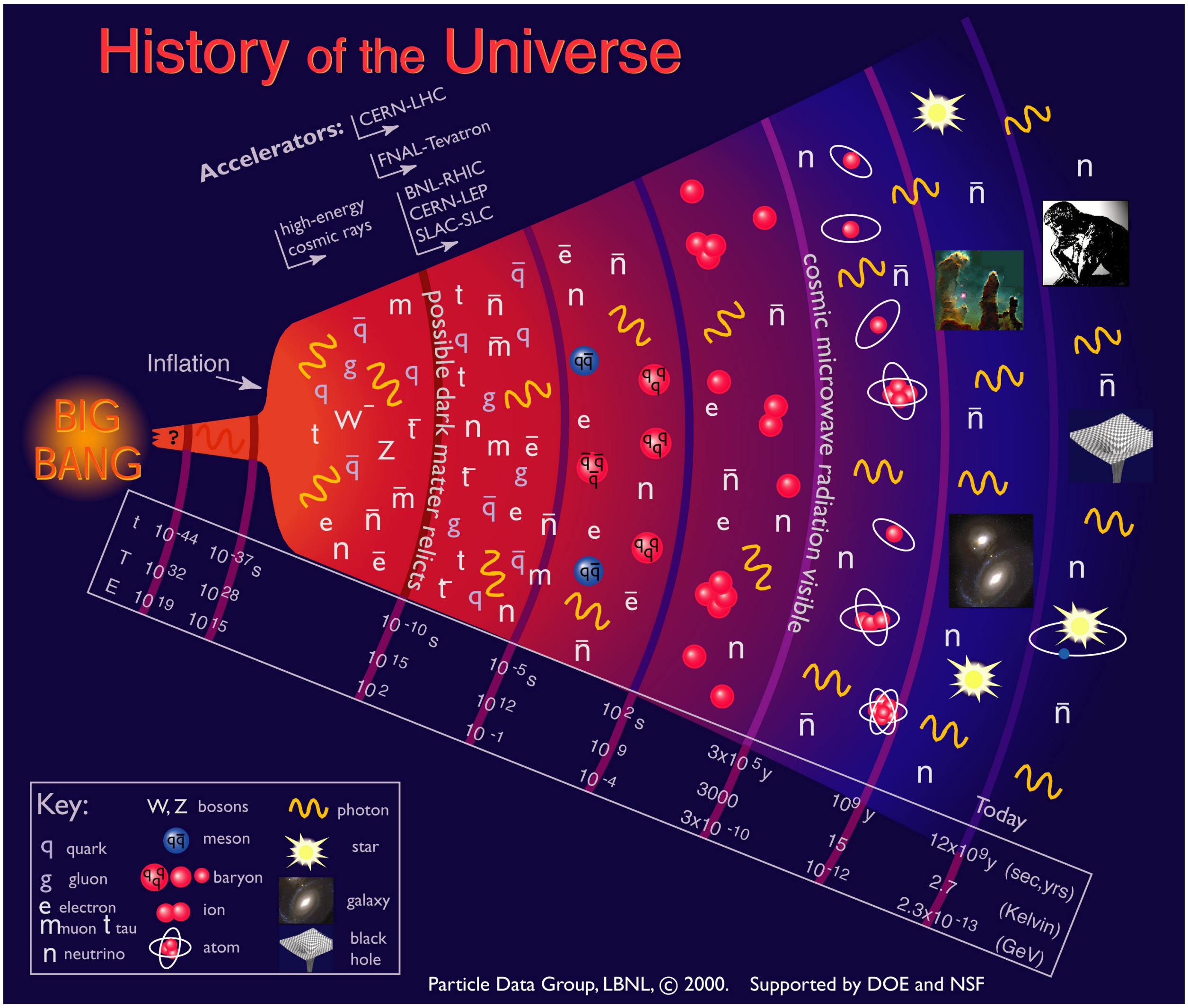
In this section we quickly summarize the important events during the evolution of our universe from its beginning until today. While this section contains nothing new, it provides a concise time-line of our universe and a summary of everything we have learned.

Event	Time	Redshift	Temperature
Planck era	$10^{-43}s$	?	$2 \times 10^{18}GeV$
GUT scale	$10^{-40}s$	?	$10^{16}GeV$
Inflation	unclear: $10^{-38} - 10^{-14}s$?	?	-
Baryogenesis	?	?	?
EW phase transition	$2 \times 10^{-11}s$	10^{15}	$100GeV$
QCD phase transition	$2 \times 10^{-5}s$	10^{12}	$150MeV$
Dark matter freeze-out	?	?	?
Neutrino decoupling	$1s$	6×10^9	$1MeV$
Electron-positron annihilation	$6s$	2×10^9	$500keV$
Big bang nucleosynthesis	$3min$	4×10^8	$100keV$
Matter-radiation equality	$60 \times 10^3 yrs$	3400	$.75eV$
Recombination	$260 - 380 \times 10^3 yrs$	1100-1400	$.26 - .33eV$
CMB	$380 \times 10^3 yrs$	1100	$.26eV$
Reionization (first stars)	$200 \times 10^6 yrs$	19	$4.7meV$
Accelerated expansion starts	$7.6 \times 10^9 yrs$	.65	$.4meV$
Formation of solar system	$9.2 \times 10^9 yrs$	.42	$.34meV$
Dark energy-matter equality	$10.2 \times 10^9 yrs$	.31	$.31meV$
Today	$13.8 \times 10^9 yrs$	0	$.24meV$

The beginning (Planck era): General relativity inevitably breaks down near the Planck scale. At this point we need a UV complete theory of quantum gravity. Our best contender, string theory, is currently not well enough understood to understand a space-like singularity like the big bang. Even if we would get a theoretical handle on such a singularity, it would be very hard to test this theory since inflation is very successful at erasing any information about the universe before inflation started.

GUT scale: The interaction strengths of the strong, weak and electromagnetic forces are functions of the energy scale. At the grand unified theory (GUT) energy


History of the Universe
Spiral Galaxy
WE ARE HERE
Heavy Atom
Black Hole
Heavy Star
Protogalaxy
Hydrogen Atom
Helium Atom
Helium Nucleus
The Desert
Big Bang
MATTER DOMINATED ERA
RADIATION DOMINATED ERA
PRESENT
15 BILLION YEARS
5 BILLION YEARS
First Supernovae
Galaxy and Star Formation
1 BILLION YEARS
THE UNIVERSE BECOMES TRANSPARENT
Decoupling of Matter-Radiation
Formation of Atoms
10^{13} sec.
PHOTON EPOCH
Nucleosynthesis of Helium
10^{2} sec.
Disappearance of Positrons
LEPTON EPOCH
Confinement of Quarks
Formation of Protons, Neutrons
Disappearance of Antiquarks
10^{-10} sec.
ELECTROWEAK EPOCH
QUARK EPOCH
Asymmetry $Q - \bar{Q}$, $L - \bar{L}$
Magnetic Monopoles?
10^{-34} sec.
Cosmic Inflation?
GRAND UNIFICATION EPOCH
10^{-43} sec.
QUANTUM GRAVITY EPOCH
microcosm CERN



scale of $10^{16} GeV$ all three are almost the same. Many people believe that the three forces get unified to a single grand unified force, since it is non-trivial that three lines intersect in a point. Breaking of this single force into the three forces we observe can lead to relics like magnetic monopoles that could overclose the universe ($\Omega \gg 1$). A period of inflation with a reheating temperature below the GUT scale would solve this problem.

Inflation: Inflation is a period of exponential expansion of our universe. The lower bound on the expansion of the scale factor during (high scale) inflation is 50-60 e-folds, i.e., $a_f/a_i \geq e^{50} - e^{60}$. Such a period solves the horizon and flatness problem but more interestingly it provides the inhomogeneities needed to explain the structure in our universe. The inflaton undergoes quantum fluctuations that get stretched during the rapid expansion and after inflation get converted into small mass density inhomogeneities. These inhomogeneities are then being enhanced due to the gravitational attraction. So, slightly denser regions will become galaxies and galaxy clusters while less dense region will become emptier voids. Thus, quantum fluctuations during inflation provide the seeds for our galaxies!

Baryogenesis: There is an asymmetry between baryons and anti-baryons (and more general between particles and anti-particles) that cannot be explained by the standard model of particle physics. Thus, at energies above $10TeV$ there must be some new physics that generates this asymmetry. While there are many different theoretical ideas, there is no experimental test of any of these. So, we cannot associate a time to baryogenesis. Since the observed universe is neutral under the electric charge, there must be a similar asymmetry between electrons and positrons so that after their annihilation we are left with one electron for each proton. The ample interactions in our early universe ensure that any baryon asymmetry leads automatically to a lepton asymmetry and vice versa. Therefore, it is also possible that our early universe underwent leptogenesis instead of baryogenesis or both.

Electroweak-phase transition: During this phase transition particles get their mass due to the so-called Higgs effect. Once the standard model particles are massive, they start to drop out of equilibrium whenever the temperature of the universe (i.e., the thermal bath) becomes smaller than their mass. Then the particles start to annihilate with their anti-particles and their number densities decrease exponentially. The remaining matter in our observed universe is due to the matter-anti-matter asymmetry mentioned above.

QCD phase transition: The strong force is weaker at higher energies (temperatures) and becomes stronger and stronger during the cooling of the universe. Around $150MeV$ the strong force is so strong that free gluons and quarks cannot exist anymore and all the quarks are bound into so called baryons and mesons. These are bound states that are neutral under the strong force. The lightest baryons are the familiar proton and neutron. There are also heavier baryons and mesons that can be lighter than the proton and neutron but all of these are unstable and quickly decay. So, a little bit after the QCD phase transition we are left with essentially only protons and neutrons that are the building blocks for the atomic nuclei.

Dark matter freeze-out: If we assume that the unknown dark matter (DM) is a very weakly interacting, massive particle that was initially in equilibrium with the standard model particles, then it should freeze-out around or before the neutrino decoupling to give the correct relic abundance that we observe today, i.e., to provide a contribution to the energy density today that is roughly five times as large as the contribution of the regular matter (RM) ($\Omega_{DM} \approx .25 \approx 5\Omega_{RM}$).

Neutrino decoupling: Around $1MeV$ the weak interaction becomes so weak that particles that are only charged under the weak force, i.e., the neutrinos, decouple from the thermal plasma. These neutrinos, similarly to the photons in the CMB, give rise to a cosmic neutrino background that is slightly colder than the CMB and very difficult to observe directly. At the time of decoupling the three neutrinos are still relativistic and during the cooling of the universe they become non-relativistic whenever their temperature becomes smaller than their respective mass. Note however that this does not mean that their number density will decay exponentially, since the neutrinos are decoupled from themselves so that they cannot annihilate with each other.

Electron-positron annihilation: Around $T \sim m_e \approx 511keV$ the electrons and positrons become non-relativistic and transfer their energy and entropy into the photons only (since the neutrinos are decoupled already). This slows down the decrease in the temperature of the photons a little bit so that the photons today have a temperature that is a little bit larger than the neutrino background.

Big bang nucleosynthesis: One of the greatest successes of the big bang cosmology is that it correctly predicts the observed abundance of elements in our universe. Using nuclear physics, we can predict the amounts of different elements in the early universe and these predictions agree with what we observe, in particular besides traces of heavier elements our universe consists of 93% Hydrogen and 7% Helium. Any kind of new physics that can appear beyond the standard model is severely constrained by this success.

Recombination: Once the average energy of the photons drops below $.33eV$ the tail of high energy photons is sufficiently small to allow for neutral atoms to form. This process in which electrons and protons combine takes roughly 100,000 years and at its end the universe is filled with clouds of neutral atoms and the cosmic microwave background.

The cosmic microwave background (CMB): Once the electrons and nuclei combine into neutral atoms, the photons can stream freely until today. The observation of this cosmic microwave background does not only tell us about the universe 380,000 years after the big bang but the incredible homogeneity of the CMB also strongly motivates a phase of inflation in our very early universe. The small deviations from homogeneity in the CMB photons we observe together with their polarization provide detailed information about this period of inflation.

Reionization (first stars): The formation of the first stars leads to the release of large amounts of energy from the nuclear fusion in the stars. This energy is emitted from the stars via photons and these photons reionize the neutral atoms in the universe that are in large clouds and which will provide the fuel for future generations of stars. (Star formation should end around 10^{14} years from now, so there is still plenty of fuel out there.) The nuclear fusion in the first stars also creates the heavy elements that we observe in our universe and that were not created during big bang nucleosynthesis.

Accelerated expansion starts: The standard forms of energy density like radiation and non-relativistic matter lead to a deceleration of the expansion of our universe, i.e., $\ddot{a}(t) < 0$. This means that since the end of inflation our universe is expanding but at a decelerating rate. Due to the presence of a positive cosmological constant our universe started to expand at an accelerating rate roughly 6.2 Gyrs ago. Since matter gets diluted away during the further expansion of our universe, while the energy density due to the cosmological constant remains constant our universe is asymptotically approaching a de Sitter phase in its future.

Formation of the solar system: As a reference point, I included the age of the solar system which formed around $4.6Gyrs$ ago. Our milky way contains much older stars and its age is believed to be around $13.2Gyrs$. The presence of older stars in our vicinity is required in order to explain the abundance of heavy elements in our solar system. These heavier elements are created via nuclear fusion in the first stars and then released during supernovae.

Dark energy-matter equality: Our current universe consists of roughly 70% dark energy and 30% matter (out of which roughly 25% is dark matter and 5% is regular matter like Hydrogen and Helium). Matter gets diluted during the expansion of the universe while the energy density of the cosmological constant does not. This means that in the not too distant past, roughly $3.6Gyrs$ ago, the energy density of the universe was consisting to 50% of dark energy and to 50% of matter. Note that the accelerated expansion due to the cosmological constant did start earlier.

Today: The age of our universe is roughly $13.8Gyrs$ where the last digit can still change due to the uncertainty in the Hubble parameter. However, there are a variety of different experiments that all place mutually consistent bounds on the age of the universe so that the age of our universe is undoubtedly finite.

A Deriving the Friedmann equations from general relativity

The FRW metric in Cartesian coordinates is

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2 + g_{ij}dx^i dx^j = -dt^2 + a(t)^2\left(dx_i^2 + K\frac{(x_i dx_i)^2}{1-Kx_i^2}\right), \quad \text{(A.1)}$$

where Greek letters run over $\mu, \nu, \ldots = 0, 1, 2, 3$ and latin letters $i, j, \ldots = 1, 2, 3$. The Christoffel symbol $\Gamma^\rho_{\mu\nu}$ is given by

$$\Gamma^\rho_{\mu\nu} = \frac{1}{2}g^{\rho\sigma}\left[\frac{\partial g_{\sigma\mu}}{\partial x^\nu} + \frac{\partial g_{\sigma\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right]. \quad \text{(A.2)}$$

For the metric (A.1) we find the following non-zero components

$$\Gamma^0_{ij} = \frac{\dot{a}(t)}{a(t)}g_{ij}, \quad \text{(A.3)}$$

$$\Gamma^i_{0j} = \frac{\dot{a}(t)}{a(t)}\delta^i_j, \quad \text{(A.4)}$$

$$\Gamma^i_{jk} = \frac{Kx^i g_{jk}}{a(t)^2}. \quad \text{(A.5)}$$

From these we can calculate the Riemann curvature tensor

$$R^\mu{}_{\nu\rho\sigma} = \partial_\rho\Gamma^\mu_{\nu\sigma} - \partial_\sigma\Gamma^\mu_{\nu\rho} + \Gamma^\mu_{\alpha\rho}\Gamma^\alpha_{\nu\sigma} - \Gamma^\mu_{\alpha\sigma}\Gamma^\alpha_{\nu\rho}. \quad \text{(A.6)}$$

I will not list all non-zero components here since this is not overly illuminating and we are only interested in the Ricci curvature tensor and the Ricci scalar

$$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu}, \qquad R = g^{\mu\nu}R_{\mu\nu}. \quad \text{(A.7)}$$

The components of the Ricci tensor are

$$R_{00} = -3\frac{\ddot{a}(t)}{a(t)}, \quad \text{(A.8)}$$

$$R_{0i} = 0, \quad \text{(A.9)}$$

$$R_{ij} = \frac{\ddot{a}(t)a(t) + 2\dot{a}(t)^2 + 2K}{a(t)^2}g_{ij}, \quad \text{(A.10)}$$

where as expected the isotropy and homogeneity of our metric leads to the vanishing of the vector $R_{i0} = 0$ and forces the spacial part to be proportional to the metric $R_{ij} \propto g_{ij}$. The Ricci scalar is given by

$$R = \frac{6\left(a(t)\ddot{a}(t) + \dot{a}(t)^2 + K\right)}{a(t)^2}. \quad \text{(A.11)}$$

We recall from subsection 1.3 that the energy momentum tensor $T_{\mu\nu}$ is similarly constraint as the Ricci scalar. It can only contain two independent functions of t and its components are

$$T_{00} = \rho(t), \quad \text{(A.12)}$$

$$T_{0i} = 0, \quad \text{(A.13)}$$

$$T_{ij} = p(t)g_{ij}. \quad \text{(A.14)}$$

Now we can solve Einstein's equations

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G \, T_{\mu\nu} \,. \tag{A.15}$$

First let us look at the (00) component

$$\begin{aligned} -3\frac{\ddot{a}(t)}{a(t)} + \frac{3\left(a(t)\ddot{a}(t) + \dot{a}(t)^2 + K\right)}{a(t)^2} - \Lambda &= 8\pi G \, \rho(t) \\ \frac{3\left(\dot{a}(t)^2 + K\right)}{a(t)^2} - \Lambda &= 8\pi G \, \rho(t) \,. \end{aligned} \tag{A.16}$$

Dividing both sides by 3 leads to the first Friedmann equations as given in equation (1.24)

$$\frac{\dot{a}(t)^2 + K}{a(t)^2} - \frac{\Lambda}{3} = \frac{8\pi G}{3} \, \rho(t) \,. \tag{A.17}$$

The mixed components $(0i)$ all vanish and the pure spacial part takes the form

$$\begin{aligned} \frac{\ddot{a}(t)a(t) + 2\dot{a}(t)^2 + 2K}{a(t)^2} g_{ij} - \frac{3\left(a(t)\ddot{a}(t) + \dot{a}(t)^2 + K\right)}{a(t)^2} g_{ij} + \Lambda g_{ij} &= 8\pi G \, p(t) g_{ij} \\ \left(-2\frac{\ddot{a}(t)}{a(t)} - \frac{\dot{a}(t)^2 + K}{a(t)^2} + \Lambda\right) g_{ij} &= 8\pi G \, p(t) g_{ij} \,. \end{aligned} \tag{A.18}$$

Since the metric $g_{ij} \neq 0$ we can drop it and plug in (A.17) to get

$$-2\frac{\ddot{a}(t)}{a(t)} - \frac{8\pi G}{3} \, \rho(t) - \frac{\Lambda}{3} + \Lambda = 8\pi G \, p(t) \tag{A.19}$$

$$-2\frac{\ddot{a}(t)}{a(t)} + \frac{2}{3}\Lambda = 8\pi G \, p(t) + \frac{8\pi G}{3} \, \rho(t) \,. \tag{A.20}$$

Dividing by -2 leads to the second Friedmann equation as given in (1.25)

$$\frac{\ddot{a}(t)}{a(t)} - \frac{1}{3}\Lambda = -\frac{4\pi G}{3} \left(\rho(t) + 3p(t)\right) \,. \tag{A.21}$$

B Deriving the energy momentum tensor for a scalar field

The energy momentum tensor is defined as the variation of the action with respect to the metric $g_{\mu\nu}$. For inflation we are interested in the action of a scalar field that is given by

$$S = \int d^4x \sqrt{-g} \mathcal{L} = \int d^4x \sqrt{-g} \left(-\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \, \partial_\nu \phi - V(\phi)\right) \,. \tag{B.1}$$

Before we vary this action with respect to the metric $g_{\mu\nu}$ we recall the variations of $\sqrt{-g} = \sqrt{-\det(g_{\mu\nu})}$ and the inverse metric $g^{\mu\nu}$:

$$\delta\sqrt{-g} = -\frac{1}{2\sqrt{-g}} \delta g = \frac{1}{2} \sqrt{-g} \, g^{\mu\nu} \delta g_{\mu\nu} \,, \tag{B.2}$$

$$\delta g^{\mu\nu} = -g^{\mu\alpha} g^{\nu\beta} \delta g_{\alpha\beta} \,, \tag{B.3}$$

where in the first line we used Jacobi's formula $\delta g = g\, g^{\mu\nu}\delta g_{\mu\nu}$.

Now we can calculate the energy momentum tensor for a single scalar field

$$\begin{aligned} T^{\mu\nu} &= \frac{2}{\sqrt{-g}}\frac{\delta(\sqrt{-g}\mathcal{L})}{\delta g_{\mu\nu}} \\ &= \frac{2}{\sqrt{-g}}\left(\frac{1}{2}\sqrt{-g}\, g^{\mu\nu}\mathcal{L} + \sqrt{-g}\frac{\delta\mathcal{L}}{\delta g_{\mu\nu}}\right) \\ &= g^{\mu\nu}\left(-\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\,\partial_\beta\phi - V(\phi)\right) + g^{\mu\alpha}g^{\nu\beta}\partial_\alpha\phi\,\partial_\beta\phi \\ &= g^{\mu\nu}\left(-\frac{1}{2}\partial_\alpha\phi\,\partial^\alpha\phi - V(\phi)\right) + \partial^\mu\phi\,\partial^\nu\phi\,. \end{aligned} \tag{B.4}$$

Lowering the indices, we find

$$\begin{aligned} T_{00} &= \rho_\phi = \frac{1}{2}\partial_\alpha\phi\,\partial^\alpha\phi + V(\phi) + \dot{\phi}^2\,, \\ T_{ij} &= P_\phi g_{ij} = g_{ij}\left(-\frac{1}{2}\partial_\alpha\phi\,\partial^\alpha\phi - V(\phi)\right) + \partial_i\phi\,\partial_j\phi\,. \end{aligned} \tag{B.5}$$

Recalling the FRW metric

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2 + a(t)^2\gamma_{ij}dx^i dx^j \equiv -dt^2 + a(t)^2\left(dx_i^2 + K\frac{x_i^2 dx_i^2}{1-Kx_i^2}\right)\,, \tag{B.6}$$

we can read of the energy density and pressure for a scalar field [29]

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}\frac{(\nabla\phi)^2}{a^2} + V(\phi)\,, \tag{B.7}$$

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - \frac{1}{6}\frac{(\nabla\phi)^2}{a^2} - V(\phi)\,, \tag{B.8}$$

where $(\nabla\phi)^2 = \gamma^{ij}\partial_i\phi\,\partial_j\phi$ with γ^{ij} being the inverse of the γ_{ij} defined in equation (**??**). This is the expected result and we see that a slowly varying scalar field indeed behaves like a cosmological constant since $\rho_\phi \approx -P_\phi$.

Note that you can derive the full Einstein's equations from the action

$$S = \int d^4x\sqrt{-g}\mathcal{L} = \int d^4x\sqrt{-g}\left(\frac{1}{16\pi G}(R-2\Lambda) - \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi - V(\phi)\right)\,. \tag{B.9}$$

when using equation (B.2) and the fact that (up to a total derivative) $\delta R = R_{\mu\nu}\delta g^{\mu\nu}$.

[29] To get P_ϕ we can use $g^{ij}T_{ij} = g^{ij}g_{ij}P_\phi = 3P_\phi$.