Lecture Notes for PHYS 4202/5402 Cosmology

Yue Zhang
Carleton University

1 Our Universe is Not Empty

Photons:

Starlight makes our universe look fascinating, but ...

The most abundant photos are in fact the cosmic microwave background (CMB) – black body radiation – temperature 2.73 Kelvin.

Planck's Law: spectral radiance of black body

$$B(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/(kT)} - 1} \ . \tag{1}$$

The spectral radiance is defined as energy radiated per unit time per unit area per unit solid angle, at certain frequency ν, from the surface of the black body.

The CMB is a black body that fills our universe, without boundary. But we can pick up any surface in the universe and measure $B(\nu, T)$. In this case, it is related to the energy spectrum of the CMB photon. $B = E(dn_\gamma^{\rm CMB}/d\nu)c/(4\pi)$. As a result,

$$\frac{dn_\gamma^{\rm CMB}}{dE} = \frac{1}{\pi^2(\hbar c)^3} \frac{E^2}{e^{E/(kT)} - 1} \ , \tag{2}$$

where we used $E = h\nu$, $h = 2\pi\hbar$, k is Boltzmann constant, $T = 2.73\,{\rm K}$.

History: CMB background first discovered by Penzias and Wilson in 1964, won Nobel Prize in 1978.

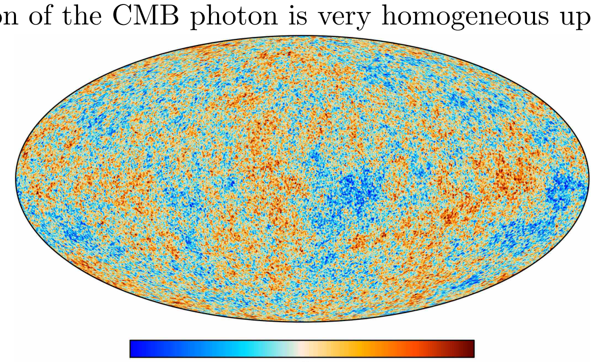
Constants of nature:

$$\begin{aligned} \hbar &= 1.05 \times 10^{-34}\,{\rm J\,s} = 0.658 \times 10^{-15}\,{\rm eV\,s} \ , \\ c &= 3 \times 10^{10}\,{\rm cm/s} \ , \\ k &= 1.38 \times 10^{-23}\,{\rm J/K} = 8.62 \times 10^{-5}\,{\rm eV/K} \end{aligned} \tag{3}$$

Number density of photon in the universe (or the space around us)

$$n_\gamma^{\rm CMB} = \frac{2\zeta(3)}{\pi^2} \left(\frac{kT}{\hbar c}\right)^3 \simeq 400\,{\rm cm}^{-3} \ . \tag{4}$$

Spatial distribution of the CMB photon is very homogeneous up to $\sim 1/10^5$.



Atoms:

Protons, neutrons, electrons that make up stars and gases.

They are the building blocks of planets, stars, galaxies, clusters of galaxies (stuff with overdensity) that we see.

Typically, a galaxy contains $\sim 10^{11}$ solar-like stars. A cluster contains $\sim 10^3$ galaxies.

A sense of length scales:

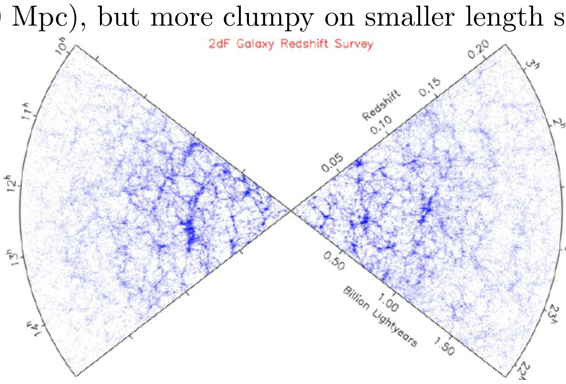
$$
\begin{aligned}
\text{Human being} &\sim 10^2\,\text{cm} \\
\text{Size of Earth} &\sim 10^4\,\text{km} = 10^9\,\text{cm} \\
\text{Size of Sun} &\sim 10^{11}\,\text{cm} \\
\text{Earth-Sun distance} &= 1\,\text{a.u.} \sim 10^{13}\,\text{cm} \\
\text{Size of Solar system} &\sim 10^{16}\,\text{cm} \\
\text{Nearest star to the sun} &\sim 1.3\,\text{pc} \sim 4\times 10^{18}\,\text{cm} \quad \text{(Proxima Centauri)} \\
\text{Distance of us from galactic center} &\sim 8\,\text{kpc} \\
\text{Size of Galaxy} &\sim 100\,\text{kpc} \sim 3\times 10^{23}\,\text{cm} \\
\text{Nearest galaxy to Milky Way} &\sim 1\,\text{Mpc} \sim 3\times 10^{24}\,\text{cm} \quad \text{(Andromeda)} \\
\text{Typical size of Galaxy cluster} &\sim 10\,\text{Mpc} \sim 3\times 10^{25}\,\text{cm} \\
\text{Large scale structure} &\sim 100\,\text{Mpc} \sim 3\times 10^{26}\,\text{cm} \\
\text{The furthest we can see} &\sim 14\,\text{Gpc} \sim 4\times 10^{28}\,\text{cm}
\end{aligned}
\tag{5}
$$

Relation among units of distance

$$1\,\text{pc} = 3.26\,\text{ly} = 2.06\times 10^5\,\text{a.u.} = 3.08\times 10^{18}\,\text{cm}\ . \tag{6}$$

Our universe is quite large.

Galaxy survey shows the spatial distribution of galaxies are also roughly homogeneous at large distances (100 Mpc), but more clumpy on smaller length scales.



Need a cosmic baryon asymmetry (matter-anti-matter asymmetry)

$$n_b \sim 10^{-10} n_\gamma \gg n_{\bar{b}}\ . \tag{7}$$

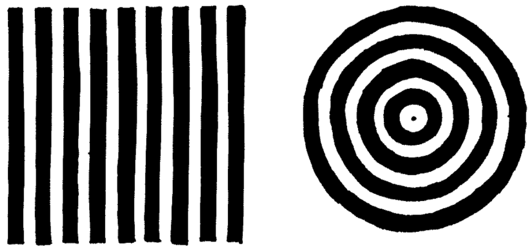
Neutrinos, dark matter, dark energy

$$n_\nu \sim n_\gamma\ . \tag{8}$$

Energy density dominated by dark side of the universe

$$\rho_{\rm DM} \sim \rho_{\rm DE} \sim 10^{-26}\,\text{kg}\,\text{m}^{-3} \sim 10^{11} M_\odot\,\text{Mpc}^{-3}\ . \tag{9}$$

$***$ Both CMB and LSS observations tell us that on large length scales, our universe appears very homogeneous and isotropic.



Left: Homogeneous but not isotropic; **Right:** isotropic but not homogeneous.

2 Age and "Size" of our Universe

Space is infinite, so is the universe.

Our universe has a finite age of ~ 14 billion years.

(Hereafter when we say "today", it can actually refer to any time during the human existence, which is much shorter than the age of universe. Our universe looks roughly the same to us right now and a few thousand years ago. As we will see in this course, the time scale of our past universe is very warped.)

Speed of light is finite.

For a far away light source, we are currently seeing its past.

We can only observe things happening at a finite distance away from us. This defines the " horizon" size. A rough estimate

$$c\tau = 3 \times 10^{10} \times 14 \times 10^9 \times 31536000 \simeq 1.3 \times 10^{28}\,\mathrm{cm}\ . \tag{10}$$

The finite age of universe also solves the paradox about why the night sky is dark.

– Assume the sun radiation power (radiated energy per unit time) is P. The earth-sun distance is $R = 1\,\mathrm{a.u.}$. The flux of light pass through unit area on the earth (e.g. your telescope) is $\Phi = P/(4\pi R^2)$.

– Next consider the universe is made of a homogeneous distribution of sun-like stars, with a number density n. In an infinitesimal volume at distance r away from us, there are $ndV = nr^2 dr d\Omega$ stars. The corresponding photons flux at earth is $d\Phi' = nPdV/(4\pi r^2) = ndrd\Omega/(4\pi)$. If the universe lives forever, integrate over the whole space. The integral is infinitely large. The paradox states stars further away are brighter than the sun because there are many of them. There are several rough edges in this calculation, when facing the reality. The key solution is the age of universe is finite, thus we only see lights from sources at finite distances. (Lights from further away are still on the way traveling to us. So the night sky tomorrow will be slightly brighter than today's? – This is actually not true.) We can only integrate r up to the horizon,

$$\Phi' = nP \int_0^{c\tau} dr\ . \tag{11}$$

Note $n \sim 10^{11}\,\mathrm{Mpc}^{-3} = 10^{-22}\,\mathrm{a.u.}^{-3}$ and $c\tau \simeq 10^{15}\,\mathrm{a.u.}$. It is clear that $\Phi' \ll \Phi$.

3 Natural Units

We are not unfamiliar with natural units.

We already had an example when defining length units. Lightyear is the distance of light travels per year. With some figure of light years, we multiply the number by speed of light to reach the distance. Effectively, this means if we keep the value of c in the back of our mind, distance is connected to time.

A further step, set $c = 1$ in our calculations. When calculating some distance but find the result has unit of time, multiply it by c to get the desired answer.

Helpful to keep the derivation simpler symbolically.

In natural units, we set $c = \hbar = k = 1$. This implies

$$[\text{distance}] = [\text{time}] = [\text{energy}]^{-1} = [\text{mass}]^{-1} = [\text{temperature}]^{-1} \ . \tag{12}$$

In the final result we supply the proper combination of constants of nature to make the unit correct.

Some useful relations

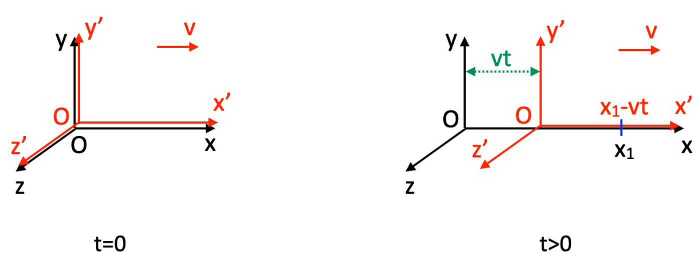
$$\begin{aligned} c &= 1 \text{ implies } 1\,\text{s} = 3 \times 10^{10}\,\text{cm} \ , \\ c &= 1 \text{ also implies } 1\,\text{eV} = 1.6 \times 10^{-19}\,\text{J} = 1.78 \times 10^{-36}\,\text{kg} \ , \\ \hbar &= 1 \text{ implies } 1\,\text{s}^{-1} = 0.658 \times 10^{-15}\,\text{eV} \ , \\ \hbar c &= 1 \text{ implies } 1\,\text{cm}^{-1} = 1.97 \times 10^{-5}\,\text{eV} \ , \\ k &= 1 \text{ implies } 1\,\text{K} = 8.6 \times 10^{-5}\,\text{eV} \ , \end{aligned} \tag{13}$$

So try to digest this concept into your blood. This way, in the upcoming discussions, you will not be surprised when I will say the proton mass is about 1 GeV, or the photon temperature at time of CMB formation is 0.1 eV etc.

4 Special Relativity

Galilean transformation: See picture below. Assume the origin of two reference frames (blue: at rest, red: moving) coincide at time $t = 0$.

$$\begin{aligned} x' &= x - vt \ , \\ y' &= y \ , \\ z' &= z \ , \\ t' &= t \ , \end{aligned} \tag{14}$$



However, this is only valid for low relative velocities ($v \ll c$).

Lorentz transformation: position of the same point particle in rest frame S, and a frame S' that moves along the $\hat{x}$ axis direction with constant velocity v,

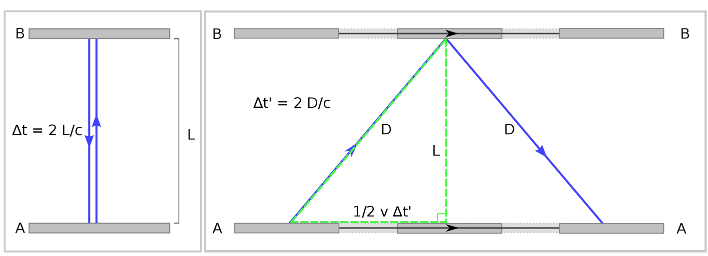
$$
\begin{aligned}
x' &= \gamma(x - vt) \ , \\
y' &= y \ , \\
z' &= z \ , \\
t' &= \gamma(t - vx/c^2) \ ,
\end{aligned} \tag{15}
$$

where $\gamma = 1/\sqrt{1 - v^2/c^2} \geq 1$ is the Lorentz boost factor. Importantly, in all frames, the speed of light is $c = 3 \times 10^{10}$ cm/s.

The Lorentz transformation returns to the Galilean transformation in the limit $v \ll c$, where $\gamma \simeq 1$.

In general, it can lead to unusual (but real) phenomena such as time dilation and length contraction.

A simple way of understanding time dilation is by considering the picture below, where we measure the speed of light by measuring the time it takes for a laser to bounce between two mirrors (separated by distance L). Imagine we build an instrument that can send and receive the laser beam. First, in the frame where the instrument is a rest respect to the mirror, the time it takes for light to travel back and forth is $\Delta t = 2L/c$. Second, in the frame where the instrument travels in horizontal direction, Lorentz transformation tells us that the distance between two mirrors remain L. Due to the moving of the instrument, the actual distance light needs to travel before it is capture by the same instrument is $\Delta t' = 2\sqrt{L^2 + (v\Delta t'/2)^2}/c$. This is an equation for $\Delta t'$. The solution is $\Delta t' = 2L/(c\sqrt{1 - v^2/c^2}) = \gamma \cdot \Delta t > \Delta t$. In a moving frame, it takes longer the complete the physical measurement.



Example: high-energy cosmic ray showering on Earth's upper atmosphere produce particles such as the muon. Muon is a heavy cousin of electron and unstable. Life time about 10^{-6} sec. In the lab frame (at rest on earth surface), the resulting muon can travel with speed very close to c. Consider a muon with γ factor equal to 1000. Distance muon can travel before it decays in the lab frame: $\gamma\beta\tau \simeq \gamma c\tau \simeq 3 \times 10^7\,\text{cm} = 300\,\text{km}$. Thickness of atmosphere about 100 km. Experimentally, we do see such high energy muons. In contrast, Newton and Galileo would think (they believe time is universal for all reference frames) the distance muon can travel is always $v\tau \sim c\tau = 30\,\text{m}$, which would be too short to penetrate the atmosphere.

Length contraction can be understood by considering the size of an object along the $\hat{x}$ direction. In the rest frame of the object, the two edges of the object is located at x_1 and x_2. Because it is at rest, it does not matter we can measure the two positions at any time. Let's define $L = x_2 - x_1$. Next, consider a moving frame boosted in the $\hat{x}$ direction and let's make another measurement. Because the object is now moving, we must make sure to measure the two edges at the same time. Lorentz transformation states $x'_i = \gamma(x_i - vt_i)$

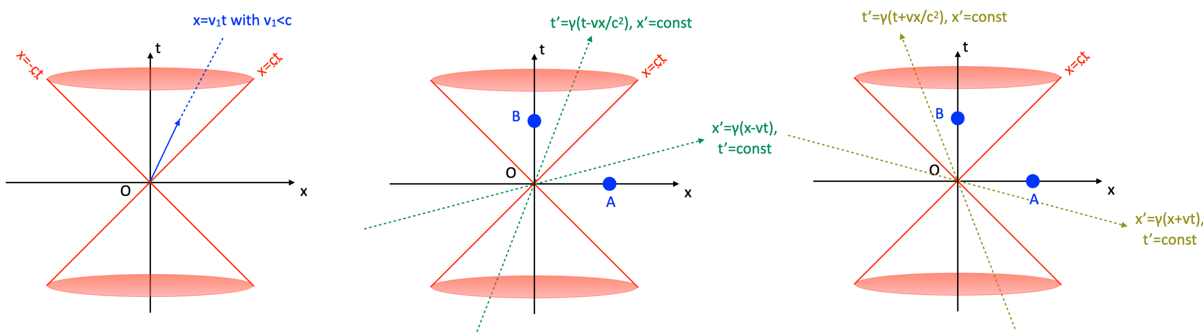
and $t'_i = \gamma(t_i - vx_i/c^2)$ where $i = 1, 2$. As mentioned, we must make sure $t'_1 = t'_2$. The second equation leads to $t_2 - t_1 = vL/c^2$. Plugging it into the first equation, we get $L' = x'_2 - x'_1 = \gamma(L - v(t_2 - t_1)) = \gamma L(1 - v^2/c^2) = L/\gamma < L$. It implies that the fast I ran toward you, the thinner I would appear.

Light cone:

In any given frame, light travels along a path such that $x = \pm ct$ (assuming it travels in the $\hat{x}$-direction). This corresponds to two diagonal lines in the x-t space. See the picture below (left figure). Generalizing this to two space dimensions the two lines rotates into a cone, called the light cone. It is hard to imagine three space dimensions, but hopefully you get the point.

A Lorentz transformation with moving observer in the $+\hat{x}$ direction with velocity v rotates the x and t axis into the x' and t' ones, as shown by the middle figure. The x', t' axis for an observer moving in opposite direction are depicted by the right figure.

Remarkably, the position of light cone stays unchanged in all reference frames.



Causality: In the above figure. A (B) lies outside (inside) the light cone from the perspective of O. Thus, O and A (O and B) are causally disconnected (connected).

Imagine two gunmen standing at points O and A are seen to shoot at each other at the same time, by an observer (1) in the x-t reference frame. A moving observer (1) in the x'-t' reference frame would see A takes the shot before O (middle figure). A third observer moving in the opposite direction to observer 2 ($v \to -v$) would reach the opposite conclusion, i.e., O takes the shot before A (right figure).

In contrast, for two gunmen located at pointed O and B. There is no ambiguity. All the observers would agree that O takes the shot before B.

Spacetime distance: OK, things change from frame to frame. But there is a quantity that remains unchanged – the spacetime distance between two points, defined as

$$(\Delta s)^2 = -(c\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \ . \tag{16}$$

In a different from moving along $\hat{x}$ direction, we have

$$\begin{aligned} \Delta x' &= \gamma(\Delta x - v\Delta t) \ , \\ \Delta y' &= \Delta y \ , \\ \Delta z' &= \Delta z \ , \\ \Delta t' &= \gamma(\Delta t - v\Delta x/c^2) \ , \end{aligned} \tag{17}$$

It is straightforward to check that

$$\begin{aligned}
(\Delta s')^2 &= -(c\Delta t')^2 - (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2 \\
&= -c^2\gamma^2(\Delta t - v\Delta x/c^2)^2 + \gamma^2(\Delta x - v\Delta t)^2 + (\Delta y)^2 + (\Delta z)^2 \\
&= -\gamma^2[c^2(\Delta t)^2 - 2v\Delta t\Delta x + v^2(\Delta x)^2/c^2] + \gamma^2[(\Delta x)^2 - 2v\Delta t\Delta x + v^2(\Delta t)^2] + (\Delta y)^2 + (\Delta z)^2 \\
&= -\gamma^2(c^2 - v^2)(\Delta t)^2 + \gamma^2(1 - v^2/c^2)(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \\
&= -(c\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \\
&= (\Delta s)^2 \ ,
\end{aligned} \tag{18}$$

where we used $\gamma = 1/\sqrt{1 - v^2/c^2}$ in the next-to-last step.

The spacetime distance is a Lorentz invariant quantity. Lorentz transformation and spacial rotation are the transformations of spacetime coordinates that keeps ds^2 and the metric invariant.

Light always travels along a path (on the light cone) in spacetime such that $\Delta s = 0$.

Note added: energy and momentum of a moving particle.
The same Lorentz transformation acting on $x^\mu = (ct, x, y, z)$ also applies to the energy-momentum vector of a particle $(E, \vec{p}c) = (E, p_x c, p_y c, p_z c)$. In the rest frame, this vector is simply $(mc^2, \vec{0})$. In the reference frame of an observer traveling in the $+\hat{x}$ direction, the energy and momentum becomes

$$\begin{pmatrix} E \\ p_x c \\ p_y c \\ p_z c \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v/c & 0 & 0 \\ -\gamma v/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} mc^2 \\ 0 \\ 0 \\ 0 \end{pmatrix} \ . \tag{19}$$

More explicitly, we have

$$E = \gamma mc^2, \quad p_x = -\gamma v m = -Ev/c^2 \ . \tag{20}$$

The minus sign is because the observer travels in $+\hat{x}$ direction, so the particle has to travel in the opposite direction.
It is straightforward to verify the on-shell relation (note $|\vec{p}| = \sqrt{p_x^2 + p_y^2 + p_z^2}$)

$$E = \sqrt{m^2c^4 + |\vec{p}|^2c^2} = \sqrt{m^2 + |\vec{p}|^2} \ . \tag{21}$$

(Last step natural unit). If a massive particle travels in the lab frame with energy E. The boost factor is simply calculated by

$$\gamma = E/(mc^2) = E/m \ . \tag{22}$$

(Last step natural unit). Correspondingly, if the particle is unstable, its lifetime becomes longer compared to rest frame by a factor of γ.

5 Metric and Space Geometry

Remarkably, in different inertial reference frames, the way to compute the spacetime distance is the same and they give the same result. This implies that there is something invariant regardless of the reference frame of the observer. The "something" here is the geometry of space and time. Let's sharpen this important point.

Consider an infinitesimal displacement, the spacetime distance can be written as

$$\begin{aligned} ds^2 &= -c^2dt^2 + dx^2 + dy^2 + dz^2 \\ &= \begin{pmatrix} cdt & dx & dy & dz \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} cdt \\ dx \\ dy \\ dz \end{pmatrix} . \end{aligned} \tag{23}$$

The square matrix in the middle is called the metric tensor, often denoted by $g_{\mu\nu}$ where $(\mu, \nu = 0, 1, 2, 3)$. The same metric applies to all the inertial reference frames (that travel with respect to each other with constant velocity).

Metric will play the key role in the discussions of the upcoming chapters.

Euclidean geometry. Let's focus on the spatial part of the metric for now and discuss the geometry of space. We first consider the case of flat space, which describes Euclidean geometry. The spatial distance is given by

$$d\ell^2 = dx^2 + dy^2 + dz^2 = dr^2 + r^2d\theta^2 + r^2 \sin^2\theta d\phi^2 . \tag{24}$$

In the second step we go to the spherical coordinates, where θ and ϕ are the polar and azimuthal angles. Clearly, such a coordinate system is both homogeneous and isotropic. There is no no preferred direction, i.e., we can choose a different $\hat{z}$ axis where the angular coordinates are θ' and ϕ' but $d\ell^2$ will still take the above form but in terms of θ' and ϕ'. We are also allowed to choose the origin of axis at any point in the same.

Such a metric is very relevant for understanding cosmology because our universe is observed to be homogeneous and isotropic at large length scales.

Flat geometry is not the unique choice for our universe. More generally, space can also be described by non-Euclidean geometries that are still homogeneous and isotropic. There are two possibilities, spherical or hyperbolic.

Spherical geometry describes a three-dimensional world living on the surface of a four-dimensional sphere with radius R. (It is perhaps easier to imagine if the dimension is lowered by one, e.g., surface of a basketball.) (It has to be a sphere otherwise the metric will not be isotropic.) There are four spatial coordinates to describe the space (x_1, x_2, x_3, x_4), but they satisfy the relations $x_1^2 + x_2^2 + x_3^2 + x_4^2 = R^2$. We can go to the spherical coordinates

$$\begin{aligned} x_1 &= R\cos\chi , \\ x_2 &= R\sin\chi\cos\theta , \\ x_3 &= R\sin\chi\sin\theta\cos\phi , \\ x_4 &= R\sin\chi\sin\theta\sin\phi . \end{aligned} \tag{25}$$

Remember we are one dimensional higher than the regular case. The latter is used to describe the whole space, but now we focus on the sphere. The infinitesimal spatial distance can now be written as

$$d\ell^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 = R^2 \left[d\chi^2 + \sin^2\chi \left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] . \tag{26}$$

Let us next define $r = R\sin\chi$, and rewrite $d\ell^2$ in terms of coordinates r, θ, ϕ,

$$d\ell^2 = \frac{dr^2}{1 - r^2/R^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right) \ . \tag{27}$$

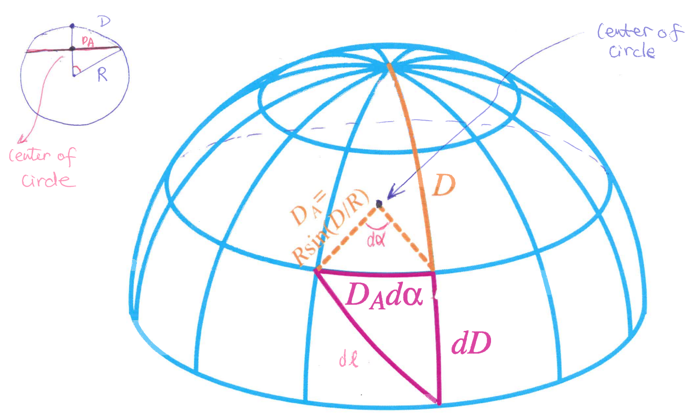
Note added: understanding the meaning of R.
Here is an intuitive example to understand the meaning of R in curved space. Let's think of a 2D sphere which is the surface of a 3D ball with radius R (indeed the surface of a basketball). We can show that the curvature of the sphere is positive. The idea is to calculate the infinitesimal spatial distance on the sphere, as shown in the picture below,

$$d\ell^2 = dD^2 + dD'^2 \ , \tag{28}$$

where dD and dD' are along the longitude and latitude directions, respectively. It is straightforward to show that $dD' = D_A d\alpha$ and $D_A = D\sin(D/R)$. As a result, the spatial distance can be written as

$$d\ell^2 = \frac{dD_A^2}{1 - \frac{D_A^2}{R^2}} + D_A^2 d\alpha^2 \ , \tag{29}$$

where we have used that $dD_A/dD = \cos(D/R) = \sqrt{1 - D_A^2/D^2}$. If we generalize the above result for a 3D sphere, then the angular measure should become $d\alpha^2 \to d\theta^2 + \sin^2\theta d\phi^2$. Comparing Eqs. (32) and (29), and renaming $D_A \to r$, we find R indeed corresponds to the radius of the sphere.



Clearly, on the surface of a sphere, the space is homogeneous and isotropic.

Hyperbolic geometry describe the surface of a hyperbolic surface. It is perhaps less familiar, but the four spatial coordinates satisfy $x_1^2 - x_2^2 - x_3^2 - x_4^2 = R^2$. Similar to the spherical coordinates, we define

$$\begin{aligned} x_1 &= R\cosh\chi \ , \\ x_2 &= R\sinh\chi\cos\theta \ , \\ x_3 &= R\sinh\chi\sin\theta\cos\phi \ , \\ x_4 &= R\sinh\chi\sin\theta\sin\phi \ . \end{aligned} \tag{30}$$

Note, $\cosh\chi = \cos(i\chi)$, $\sinh\chi = -i\sin(i\chi)$, and $\cosh^2\chi - \sinh^2\chi = 1$, and $\frac{d}{d\chi}\sinh\chi =$

$\cosh\chi$, $\frac{d}{d\chi}\cosh\chi = \sinh\chi$. The infinitesimal spatial distance on a hyperbolic surface is

$$d\ell^2 = -dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 = R^2\left[d\chi^2 + \sinh^2\chi\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] \ . \tag{31}$$

Let us next define $r = R\sinh\chi$, and rewrite $d\ell^2$ in terms of coordinates r, θ, ϕ,

$$d\ell^2 = \frac{dr^2}{1 + r^2/R^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right) \ . \tag{32}$$

Comparing Eqs. (24), (32) and (5), only the rr elements of metric are different in the three cases. We can write them in a unified way

$$d\ell^2 = \frac{dr^2}{1 - \kappa r^2/R^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right) \ , \tag{33}$$

where the parameter κ can have three possible values

$$\kappa = \begin{cases} +1, & \text{spherical geometry} \\ 0, & \text{Euclidean geometry} \\ -1, & \text{hyperbolic geometry} \end{cases} \tag{34}$$

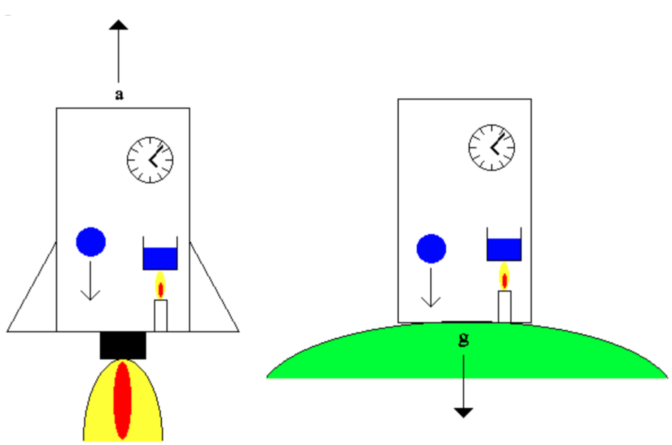
OK. Enough geometry for now.

6 Equivalence Principle

The forms of the metric discussed in the previous section are still very special cases. More generally, each matrix element of $g_{\mu\nu}$ could be a function of t and $\vec{r}$. When spacetime is not flat (or Minkowskian), special relativity is no longer useful. We enter the realm of general relativity (GR).

Our universe is an example of the latter situation. Therefore, any cosmology course cannot avoid mentioning GR. However, since this is not a GR course, we shall not go through the detailed formalism of GR. It is beyond the scope of the first cosmology course. Instead, we will discuss the essence of GR and motivate it by taking a handwaving approach.

To gain some useful intuitions, let's first consider a concrete example, as shown by the picture below. In the left plot, image you are inside a rocket that is accelerating upward with $a = 9.8\,\mathrm{m/s^2}$, whereas in the right plot you are simply standing on the surface of the earth and feels a gravitational acceleration $g = 9.8\,\mathrm{m/s^2}$.



The equivalence principle states that in both cases, the laws of physics you observe should be identical. For example, if you release an object and let it free fall, you will observe it accelerates downwards with $9.8\,\mathrm{m/s^2}$. This is not a surprising result.

Next, consider another experiment where you hold a flashlight and shine the light in the horizontal direction. In the first case, because the rocket and you accelerate upward, but the beam of light does not after it is released (and the speed of light is finite), the light beam will appear to bend downward from your point of view. In the second case, when we stand on the surface of earth, the equivalence principle dictates that we should also expect the light to bend. This is a bit counter intuitive because Newton told us that the gravitational force the earth exerts on an object is proportional to its mass but photon (light) has no mass. The reality is light indeed bends. Einstein tell us that the existence of the massive Earth causes the space to be curved, so does the accelerating rocket. The path light chooses to take is the minimal distance between the two points although it does not look like a straight line.

One may try to argue against the above reasoning by assuming that the photon mass is nonzero but very very tiny (such that it passes all the tests of $1/r$ law for the Coulomb interaction). This way, the gravitational potential would also be very very tiny. However, it still fails to produce the correct result (see discussion of tests of GR in the box below).

The above thought experiment leads to the birth of general relativity (GR).

The essence of general relativity:

- Distribution of stuff (mass, energy densities, etc) dictates the spacetime metric tensor. The latter is solution to the Einstein equation.

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \ . \tag{35}$$

  Left-hand side, geometry of spacetime. $R_{\mu\nu}$ is Ricci tensor and fully determined by the metric. Λ is a parameter, often called the cosmological constant.

$$\begin{aligned} R &= g_{\mu\nu} R^{\mu\nu} \ , & & \\ R_{\mu\nu} &\sim \partial\Gamma + \Gamma^2, & & [R_{\mu\nu} = \partial_\lambda \Gamma^\lambda{}_{\mu\nu} - \partial_\mu \gamma^\lambda{}_{\nu\lambda} + \Gamma^\lambda{}_{\mu\nu}\Gamma^\sigma{}_{\lambda\sigma} - \Gamma^\lambda{}_{\mu\sigma}\Gamma^\sigma{}_{\lambda\nu}] \ , \\ \Gamma &\sim \partial g_{\mu\nu}, & & [\Gamma^\lambda{}_{\mu\nu} = \frac{1}{2} g^{\sigma\lambda}(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu})] \ . \end{aligned} \tag{36}$$

  Right-hand side, source term of gravitation (energy-momentum tensor).

  In principle, one can solve this equation for the spacetime metric, once the source term is known.

- The motion of a point-like probing object in the background of a spacetime metric is described by the geodesic equation.

$$\frac{d^2 x^\mu}{d\tau^2} = \Gamma^\mu{}_{\nu\lambda} \frac{dx^\nu}{d\tau} \frac{dx^\lambda}{d\tau} \ . \tag{37}$$

  Of course, one should be aware there is a back reaction if the universe is filled with these point-like objects, which serve as the source term in Eq. (35).

Newtonian Limit of GR. We could work out some math details. Consider the geodesic equation of a slowly-moving point mass

$$\frac{d^2x^\beta}{d\tau^2} = \Gamma^\beta{}_{\alpha\nu}\frac{dx^\alpha}{d\tau}\frac{dx^\nu}{d\tau} \ . \tag{38}$$

where τ measures the time on the clock that travels together with the point-like object (often called the proper time), and $\Gamma^\beta{}_{\alpha\nu}$ is the Christoffel symbols, defined as

$$\Gamma^\beta{}_{\alpha\nu} = \frac{1}{2}g^{\mu\beta}(\partial_\alpha g_{\mu\nu} + \partial_\nu g_{\mu\alpha} - \partial_\mu g_{\nu\alpha}) \ . \tag{39}$$

For object in slow motion ($v \ll c$) we have $t \sim \tau$, and $dx^\alpha/d\tau \simeq (1, 0, 0, 0)$. As a result, the acceleration is

$$a^j = \frac{d^2x^j}{dt^2} \simeq \Gamma^j{}_{00} \ , \tag{40}$$

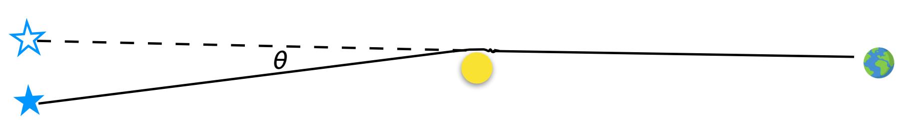
where j is the spatial index.
Let us consider the specific case of moving near the surface of the Earth. The presence of the Earth impacts the spacetime by modifying the metric tensor. The spacetime distance now takes the special form

$$ds^2 = -\left(1 - \frac{2GM}{r}\right)dt^2 + \frac{dr^2}{1 - \frac{2GM}{r}} + r^2d\theta^2 + r^2\sin^2\theta d\phi^2 \ . \tag{41}$$

This is called the Schwarzschild metric. It is time independent. (The same metric also describe the spacetime around a black hole.) The corresponding acceleration of slowly moving particle is (along the radial direction)

$$a \simeq \Gamma^r{}_{00} = \frac{1}{2}g^{rr}(0 + 0 - \partial_r g_{00}) = -\frac{1}{2}\left(1 - \frac{2GM}{r}\right)^{-1}\frac{2GM}{r^2} \simeq -\frac{GM}{r^2} \ , \tag{42}$$

where in the last step, we have use the weak gravity approximation $1 - 2GM/r \simeq 1$ for $r = R_\oplus$. Clearly, the result correctly produces the Newton's law of universal gravitation.



One can use this acceleration to calculate how much the trajectory of photon is bent by the Sun. We make an estimate here. Consider the above picture where a light travels past the edge of Sun in the $\hat{x}$ direction. The time light spends in the neighbourhood of Sun is roughly $t \simeq 2R_\odot/c$, where c is the speed of light. During this time, the light responds to the Sun's gravity and accelerates in the $\hat{y}$ (radial) direction. It gains a velocity in the $\hat{y}$ direction, $v_y \simeq at \simeq 2GM/(R_\odot c)$. The deflection angle is then $\theta \simeq v_y/c \simeq 2GM/(R_\odot c^2)$. For solar mass and radius, $\theta \sim 0.84$ arcsec.

This is actually a wrong result. See the next box.

Test of GR: deflection of light by the Sun
There is a subtlety in above discussions for massless particle. Indeed, time dilation dictates that $dt = \gamma d\tau$, where $\gamma \to \infty$ for massless particle traveling at speed of light. For infinitesimal dt, we have $d\tau = 0$, thus the geodesic equation Eq. (38) is not well defined for a photon. The trick here is to introduce an affine parameter λ to encode the path of photon such that $d\lambda = \infty d\tau > 0$. There is no need to find out the infinity parameter. For a photon traveling in the $\hat{r}$ direction, we can define $dr = Ad\lambda$. Using $ds^2 = 0$ and weak gravity approximation $g_{rr} \sim 1$, we have

$$dt \simeq dr = Ad\lambda \ . \tag{43}$$

Next, consider the geodesic equation where the differentiations are with respect to λ instead of τ, the "acceleration" in the r direction is

$$\frac{d^2r}{d\lambda^2} = \Gamma^r_{00}\left(\frac{dt}{d\lambda}\right)^2 + (\Gamma^r_{0r} + \Gamma^r_{r0})\left(\frac{dt}{d\lambda}\right)\left(\frac{dr}{d\lambda}\right) + \Gamma^r_{rr}\left(\frac{dr}{d\lambda}\right)^2 \ . \tag{44}$$

Using

$$\Gamma^r_{00} \simeq \Gamma^r_{rr} \simeq -\frac{GM}{r^2} \ , \Gamma^r_{0r} = \Gamma^r_{r0} = 0 \ , \tag{45}$$

we get

$$\frac{d^2r}{d\lambda^2} = -\frac{2GM}{r^2}A^2 \quad \Rightarrow \ , \tag{46}$$

which finally leads to the radial acceleration

$$a \equiv \frac{d^2r}{dr^2} = -\frac{2GM}{r^2} \ . \tag{47}$$

This result is twice the value found in Eq. (42). Crucially, the last term of Eq. (44) makes an extra contribution that doubles the acceleration.
As pointed out by Einstein, $\theta \simeq 1.75$ arcsec, i.e., the trajectory of light will be deflected by twice of the angle predicted using Newton's theory. This serves as an important test of GR and has been verified by the Eddington experiment, see https://en.wikipedia.org/wiki/Eddington_experiment.

7 Our Universe is Expanding

An important starting point of cosmology is to find the correct metric for our universe. To make progress, we need an important physics input here. Experimentally, cosmological observations have found that our universe is expanding.

The Hubble's law is an observation that in our universe that galaxies are moving away from the Earth at speeds proportional to their distance. Mathematically,

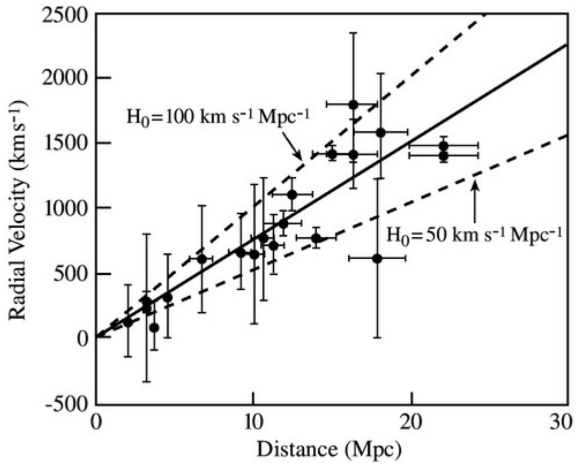
$$v = H_0 D \ , \tag{48}$$

where the proportional coefficient is introduced as the Hubble constant. The value of H_0 is

$$H_0 \simeq 70\,\mathrm{km/s/Mpc} \ . \tag{49}$$

This law only works for objects in the universe that are very far away from us (at larger than Mpc distances). Experimentally, the speed of a galaxy moving away from us can be inferred from the redshift of photon spectrum (analogy of the Doppler effect), whereas the distance can be inferred from the luminosity. We need some standard candles (such as cepheid, supernova).

From these observations, we can only know the universe has been expanding since the birth of the first stars.



From Eq. (48), one could raise a subtle question. For very large D, the speed of galaxies moving away form us can be larger than the speed of light – is this OK? The short answer is, there is nothing wrong with it. It certainly does not violate any causality. As this course goes along, we will understand it better.

An exercise employing natural unit shows

$$H_0 \simeq (4\,\mathrm{Gpc})^{-1} \sim (13\,\text{billion years})^{-1} \ . \tag{50}$$

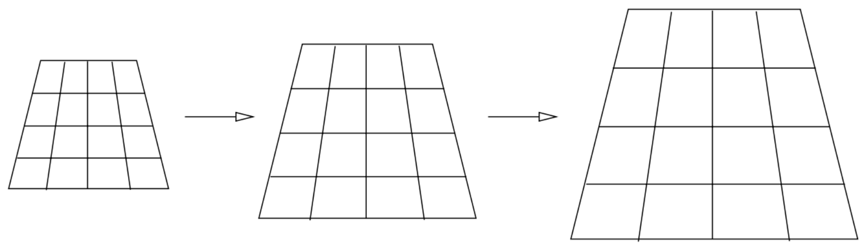
Remarkably, it roughly produces the size and lifetime of our universe. This is not a coincidence.

8 FRW Metric

So we need a metric to describe our universe, that is expanding with time. The Cosmological Principle postulates that at each point in the universe there is an observer (must be free falling, not traveling with a spaceship) to whom the universe appears isotropic and homogeneous. Each observer belongs to a special inertial reference frame. All observers share the same clock. The spatial part of the metric must satisfy one of the three geometries discussed in Section 5. The Friedmann-Robertson-Walker tensor for such an observer takes the following form, in terms of spacetime distance

$$\begin{aligned} ds^2 &= -dt^2 + a(t)^2 d\ell^2 \\ &= -dt^2 + a(t)^2 \left[\frac{dr^2}{1-\kappa r^2/R^2} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \right] \ , \end{aligned} \tag{51}$$

where r, θ, ϕ are called comoving coordinates. The infinitesimal comoving distance $d\ell$, whose square is equal to the quantity inside the bracket, is time independent. On top of that we introduce a scale factor $a(t)$. The physical (proper) spatial distance between two points on the grid is then given by $a(t) \int d\ell$. The expansion of the universe is understood as the following. As a moment t, we can take a snapshot of our universe and mark up certain points in the space. Then we let the time pass by. At a later moment $t' > t$, we take another snapshot, we will find the marked points are separated further from each other. The increase in the distances is described by the scale factor, such that $a(t') > a(t)$. It is helpful to look at the following picture.



Without loss of generality, we can set $a(t_0) = 1$, where t_0 is the age of universe today. In this case, the comoving coordinates coincides with the physical spatial coordinates if we take a snapshot of our universe at time t_0.

We also emphasize that the FRW metric only works over very large distances in the universe. It does not work if you think of physics near the earth, or within the Milky Way galaxy, where space (and the source of gravity) is clearly not homogeneous or isotropic.

The FRW metric leads to the following important concepts.

- **The comoving distance.** As discussed above, when talking about the comoving distance, we take out the scale factor a. Consider two points A and B sitting on the "spacetime grid" (free-falling observers), we can put A at the origin of comoving coordinates. We have the freedom to do this. Then B will be along the r direction with respect to A. Let's say it is located at $r = r_B$. Their comoving distance is then calculated with a radial integral

$$\ell_{AB} = \int_0^{r_B} \frac{dr}{\sqrt{1 - \kappa r^2/R^2}} = \begin{cases} R \sin^{-1}(r_B/R), & \kappa = 1 \\ r_B, & \kappa = 0 \\ R \sinh^{-1}(r_B/R), & \kappa = -1 \end{cases} \tag{52}$$

  ℓ_{AB} is time independent.

- **The physical distance.** To find the physical (or proper) distance between A and B at a given time t, we multiply the comoving distance with the corresponding scale factor $a(t)$,

$$\ell_{AB,\,\text{physical}}(t) = a(t)\ell_{AB} \ . \tag{53}$$

- **The Hubble parameter.** With a time dependent scale factor, we can define the Hubble parameter,

$$H = \frac{\dot{a}}{a} = \frac{1}{a}\frac{da}{dt} \ . \tag{54}$$

  It measures the expansion rate of the universe. The Hubble parameter is not a constant of time. It evolves along with a throughout the history of universe. The value reported in Eq. (49) is the expansion rate of the universe today. Its value does not have to be (in fact, it is not) the same in the history of universe.

- **Relative motion between far-away objects.** For the two points A and B sitting on the grid (at rest within the comoving coordinates), they are still moving away from (or toward if $H < 0$) each other due to the expansion of the universe,

$$\frac{d\ell_{AB,\,\mathrm{physical}}}{dt} = \frac{d(a\ell_{AB})}{dt} = H\ell_{AB,\,\mathrm{physical}} \ . \tag{55}$$

This is exact why the FRW metric is motivated by the Hubble's law (Eq. (48)). On top of the comoving motion, A or B could have a peculiar velocity with respect to the grid, e.g., A could be orbiting around a galaxy or planet near that grid point which exerts local gravitational forces.

- **Dilation of time.** Next, imagine A is a source (astrophysical body like star) that can send photons to B. Let's imagine a simplified case where A sends photons at a constant frequency and all photons have the same wavelength λ. These photons travel from A to B as the universe is expanding. Two things will happen.

  First, the separation of two neighboring pulses become larger and larger as they travel further away from A. This is due to the expansion of space between the pulses, and the pulses left A earlier have experienced longer period of expansion. They become more and more sparse. Let's try to be a bit quantitative here. Imagine at time t_1, A sends a photon that is received by B at a later time t_2. We know light travels along a path in spacetime where the $ds = 0$, i.e., $dt = a(t)dr/\sqrt{1-\kappa r^2/R^2}$. Integrating from A to B, we get

$$\int_{t_1}^{t_2} \frac{dt}{a(t)} = \int_0^{r_B} \frac{dr}{\sqrt{1-\kappa r^2/R^2}} \ . \tag{56}$$

  Next, at a slightly later time $t_1 + \delta t_1$, A sends another photon to B which arrives at time $t_2 + \delta t_2$. Similarly we can write,

$$\int_{t_1+\delta t_1}^{t_2+\delta t_2} \frac{dt}{a(t)} = \int_0^{r_B} \frac{dr}{\sqrt{1-\kappa r^2/R^2}} \ . \tag{57}$$

  The right-hand sides of Eqs. (56) and (57) are identical, which implies

$$\int_{t_1+\delta t_1}^{t_2+\delta t_2} \frac{dt}{a(t)} - \int_{t_1}^{t_2} \frac{dt}{a(t)} = 0, \quad \Rightarrow \quad \int_{t_2}^{t_2+\delta t_2} \frac{dt}{a(t)} - \int_{t_1}^{t_1+\delta t_1} \frac{dt}{a(t)} = 0 \ . \tag{58}$$

  For very small δt_1 and δt_2 (A takes a quick action sending the two photons one after another), we find

$$\frac{\delta t_2}{a(t_2)} = \frac{\delta t_1}{a(t_1)}, \quad \Rightarrow \quad \delta t_2 = \frac{a(t_2)}{a(t_1)}\delta t_1 \ . \tag{59}$$

  This tells that the time intervals for photon radiation at A and photon receiving at B are not the same. Because $t_1 < t_2$, we have $a(t_2)/a(t_1) > 1$. It takes longer for B to receive two neighboring photons than for A to send them.

- **Dilation of space.** Next, let's forget about B and just consider the source A sending photons at constant frequency. After every time interval Δt a photon leaves A. Consider the following three snap shots of the universe: 1) at time t, A is about the send out the first photon; 2) at time $t + \Delta t$, the first photon travels to a physics distance of d_1 from A, and the second photon is about to be sent out by A; 3) at time $t + 2\Delta t$, both the first and second photon travel a distance d_2. One must notice that during $t + \Delta t$ and $t + 2\Delta t$, the earlier distance d_1 increases to d_1' along with the expansion of the universe.

We want to show that at time $t + 2\Delta t$, photon #1 is more than twice far away from A than photon #2, i.e., $d_1' + d_2 > 2d_2$, or $d_1' > d_2$. All is due to the expansion of the universe.

At time $t + \Delta t$, we have

$$d_1 = a(t + \Delta t) \int_t^{t+\Delta t} \frac{cdt_1}{a(t_1)} \ , \tag{60}$$

where the ratio $a(t + \Delta t)/a(t_1)$ is how much the infinitesimal distance photon travels during dt_1 increases at a later time $t + \Delta t$ due to the expansion of the universe.

At time $t + 2\Delta t$, clearly we have

$$d_1' = \frac{a(t + 2\Delta t)}{a(t + \Delta t)} d_1 = a(t + 2\Delta t) \int_t^{t+\Delta t} \frac{cdt_1}{a(t_1)} \ . \tag{61}$$

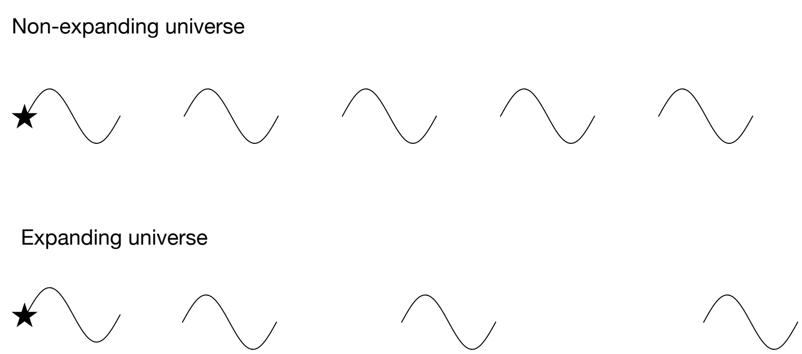
In the last step, let's make a change of integral variable by introducing $t_2 = t_1 + \Delta t$. We can write

$$d_1' = a(t + 2\Delta t) \int_{t+\Delta t}^{t+2\Delta t} \frac{cdt_2}{a(t_2 - \Delta t)} \ . \tag{62}$$

Because $a(t_2 - \Delta t) < a(t_2)$, $\Delta t > 0$ and the universe expands,

$$d_1' > a(t + 2\Delta t) \int_{t+\Delta t}^{t+2\Delta t} \frac{cdt_2}{a(t_2)} = d_2 \ . \tag{63}$$

The right-hand side is exact the distance traveled by photons during $t+\Delta t$ and $t+2\Delta t$, following the same consideration as writing down Eq. (60). This proves our goal and yields the following picture. In an expanding universe, photons leaving source A are appear more and more sparse, different from to the case of a non-expanding universe.



- **Red shift.** In cosmology, the observers are always human beings living at time t_0. Following our convention, $a(t_0) = 1$. Any light source could come from the far away past, say time t_1. It is often defined

$$1 + z = \frac{1}{a(t_1)} \ , \tag{64}$$

where z is called the redshift. Some events happening in early universe with $t_1 < t_0$ corresponds to some value of redshift $z > 0$.

Second, due to particle-wave duality, each photon also corresponds to a wave. The wavelength grows with the expansion of the universe. In term of equation, we have

$$\lambda(t_0) = (1+z)\lambda(t_1), \quad \Rightarrow \quad E(t_0) = \frac{E(t_1)}{1+z} \ . \tag{65}$$

In the second step we resort to the relation $E = hc/\lambda$ as suggested by Planck. This agrees well with our intuition so I will not prove it.

- **Luminosity distance.** Suppose we understand the source A very well. For example we have observed similar sources in nearby regions (for example in our galaxy). Then A is a standard candle, and mathematically, we know the power of its radiation, $P = dE/dt$. We would think a universal power also applies to similar sources located far away throughout the universe. Let's now consider such a far away source so that expansion of universe matters. For us to see it today, A has to radiate photons at early time t_1. At the source, the power of the standard candle is $P = dE(t_1)/dt_1$. According to Eqs. (59) and (65), we derive that when the observe receives it at t_0, the power must appear to be $P' = P/(1+z)^2$. Experimentally, what is observed is not the total power but the flux of light (power per unit area of your telescope). The flux seen by the observed will be

$$\Phi = \frac{P'}{4\pi \ell_{AB,\,\text{physical}}(t_0)^2} = \frac{P}{4\pi(1+z)^2\ell_{AB}^2} \ . \tag{66}$$

Because the observer always has the standard candle (the universal power P) in mind, we would think we measured the source A is located at distance $(1+z)\ell_{AB}$ away. This defined a quantity the luminosity distance

$$d_L = (1+z)\ell_{AB} \ , \tag{67}$$

which is a function of its redshift z.

- **Measuring H_0.** For a source located at low redshift, or $H_0(t_0 - t_1) \ll 1$, we can make the following Taylor expansion (for $t_1 \le t \le t_0$)

$$a(t) = 1 + \left.\frac{da}{dt}\right|_{t=t_0} (t - t_0) + \cdots = 1 + H_0(t-t_0) + \cdots \ , \tag{68}$$

where we have used $a(t_0) = 1$.With this expansion, we calculate the luminosity distance using Eqs. (67), (52) and (56), (assuming B is the present day observer, thus $t_2 = t_0$)

$$\begin{aligned} d_L &= (1+z)\int_{t_1}^{t_0} \frac{dt}{a(t)} \simeq (1+z)\int_{t_1}^{t_0} \frac{dt}{1+H_0(t-t_0)} \\ &\simeq (1+z)\int_{t_1}^{t_0} dt\,[1 - H_0(t-t_0)] \\ &= (1+z)(t_0-t_1)\left[1 + \frac{1}{2}H_0(t_0-t_1)\right] \end{aligned} \tag{69}$$

On the other hand, we can also find the redshift z in terms of the the above expansion,

$$z(t_1) = \frac{1}{a(t_1)} - 1 = \frac{1}{1+H_0(t_1-t_0)} - 1 \simeq H_0(t_0-t_1) \ . \tag{70}$$

The assumption that $H_0(t_0 - t_1) \ll 1$ implies $z \ll 1$. This relation allows us to solve $t_0 - t_1$ in terms of z in Eq. (71),

$$d_L = (1+z)\frac{z}{H_0}\left(1 + \frac{1}{2}z\right) \simeq \frac{z}{H_0} + \mathcal{O}(z^2) \ . \tag{71}$$

Experimentally, when observing a standard candle, we can infer its luminosity distance using Eq. (66) by measuring the flux Φ and with the knowledge of P. In addition, we can infer the redshift z by exploring the wavelength enhancement (see Eq. (65)) of typical atomic lines, or the shape of the photon spectrum (if continuous). After these two measurements, we can determine the Hubble parameter today using Eq. (71).

If we perform the above expansion to z^2 order, we will get

$$H_0 d_L = z + \frac{1}{2}(1 - q_0)z^2 + \mathcal{O}(z^3) \ , \tag{72}$$

where $q_0 \equiv -\left.\frac{d^2a}{dt^2}\right|_{t=t_0} \frac{1}{H_0^2}$ is called the deceleration parameter of the universe. This implies that if $q_0 \neq 0$ the relation between d_L and z is not linear, but parabolic at this order. We can determine q_0 by fitting the data from multiple standard candle measurements.

- **The cosmological horizon.** Horizon means boundary of the largest distance you can see. In cosmology, it corresponds to the distance light can possibly travel since the beginning of the universe (big bang), assuming the universe is transparent to the light. If you sit at the origin of axis, then there is a sphere whose radius is given by this distance. Because no particle can travel faster than the light, such a radius is also called the particle horizon. It is calculated as

$$d_H(t) = a(t)\int_0^t \frac{dt'}{a(t')} \ . \tag{73}$$

The physical meaning of the $1/a$ factor in the integrand is clear. Setting $c = 1$, dt' is the infinitesimal distance light travels during the infinitesimal time interval at an earlier time t'. This distance will experience the expansion of the universe. At a later time t, the distance becomes $a(t)dt'/a(t')$. Integrating over the history of time leads to the particle horizon. The size of the cosmological horizon today (the farthest we could possibly see) is

$$d_H(t_0) = \int_0^{t_0} \frac{dt}{a(t)} \ , \tag{74}$$

where we used $a(t_0) = 1$.

We cannot complete the integral without knowing the form of $a(t)$, which is the task of upcoming sections.

9 Stuff in Our Universe

So far, we have not derived the exact form of the scale function $a(t)$ yet. The time evolution of $a(t)$ can be solved from the Einstein equation,

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \ . \tag{75}$$

On the left-hand side of the equation, the quantities $R_{\mu\nu}$, R, $g_{\mu\nu}$ are all made of the metric tensor and depend on $a(t)$ and its time derivatives. Λ is the cosmological constant. The right-hand side is the stress-energy tensor describing all the stuff that fills our universe. Once we know what they are, we can solve the equation for $a(t)$.

The stuff in our universe are described by perfect fluids, which are completely characterized by two quantities, the energy density ρ and pressure density p. The fluid is static in the comoving frame. The stress-energy tensor takes the form

$$T^{\mu}{}_{\nu} = (\rho + p)u^{\mu}u_{\nu} + pg^{\mu}{}_{\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} , \quad T_{\mu\nu} = g_{\mu\alpha}T^{\alpha}{}_{\nu} \ , \tag{76}$$

where we used $u^{\mu} = (1, 0, 0, 0)$ and $u_{\nu} = g_{\nu\mu}u^{\mu} = (-1, 0, 0, 0)$ in the rest frame of the fluid. Because the universe is homogeneous, ρ and p are independent of the spatial position, but will depend on time.

The energy density and pressure are related to each other by a equation of state

$$p = w\rho \ . \tag{77}$$

For different kind of fluids, w can take different (constant) values. For radiation $w = 1/3$, for non-relativistic atoms $w \simeq 0$.

Now consider a piece of volume in the expanding universe, V. Inside this volume, the total energy is $E = \rho V$. Now consider the first law of thermodynamics,

$$dE = dQ - PdV \ . \tag{78}$$

Here, the heat transfer $dQ = 0$ because universe homogeneous. And in the FRW universe, $V = V_0 a^3$, where V_0 is the size of this volume today. Thus we can write

$$\begin{aligned} \frac{dE}{dt} &= -p\frac{dV}{dt} \ , \\ \Rightarrow \frac{d(\rho V_0 a^3)}{dt} &= -p\frac{d(V_0 a^3)}{dt} \ , \\ \Rightarrow \frac{d\rho}{dt}V_0 a^3 + \rho V_0 \frac{da^3}{dt} &= -pV_0\frac{da^3}{dt} \ , \\ \Rightarrow \frac{d\rho}{dt} &= -(\rho + p)\frac{1}{a^3}\frac{da^3}{dt} \ , \\ \Rightarrow \frac{d\rho}{dt} &= -3(\rho + p)\frac{1}{a}\frac{da}{dt} \ , \\ \Rightarrow \frac{d\rho}{dt} &= -3\rho(1 + w)\frac{1}{a}\frac{da}{dt} \ . \end{aligned} \tag{79}$$

From the last step, we derive how energy density of a particular species of fluid evolve with the scale factor of the universe (the continuity equation)

$$\rho(t) = \rho(t_0)a(t)^{-3(1+w)} \ . \tag{80}$$

For non-relativistic matter (e.g. atoms) $\rho \sim a^{-3}$, which clearly indicates dilution with increase of volume. For radiation $\rho \sim a^{-4}$. Because radiation has a non-zero pressure, it does work as the volume expands, resulting in ρ decreasing faster.

10 Friedmann Equation

The Friedmann equation is derived from Einstein equation, Eq. (75). It dictates the time evolution of scale factor and takes the form

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho + \Lambda}{3} - \frac{\kappa}{R^2 a^2} \ . \tag{81}$$

The left-hand side of the equation is equal to square of the Hubble parameter. How fast our universe expands depends on the energy density in our universe.

The Einstein equation also leads to the Raychaudhuri equation

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = -8\pi G p + \Lambda - \frac{\kappa}{R^2 a^2} \ . \tag{82}$$

The two equations are littered with parameters of the model such as Λ, κ and R.

Let's first consider the vacuum (setting $T_{\mu\nu} = 0$ in Einstein equation, we could think the universe is left with vacuum) as a fluid as well, and define its energy density and pressure as

$$\rho_\Lambda = \frac{\Lambda}{8\pi G}, \quad p_\Lambda = -\frac{\Lambda}{8\pi G} = -\rho_\Lambda \ , \tag{83}$$

which implies for vacuum $w = -1$. This is consistent with the continuity equation, Eq. (80), because ρ_Λ is a constant of time.

Next, we make similar definitions for the space curvature term

$$\rho_\kappa = -\frac{3\kappa}{8\pi G R^2 a^2}, \quad p_\kappa = \frac{\kappa}{8\pi G R^2 a^2} = -\frac{1}{3}\rho_\kappa \ . \tag{84}$$

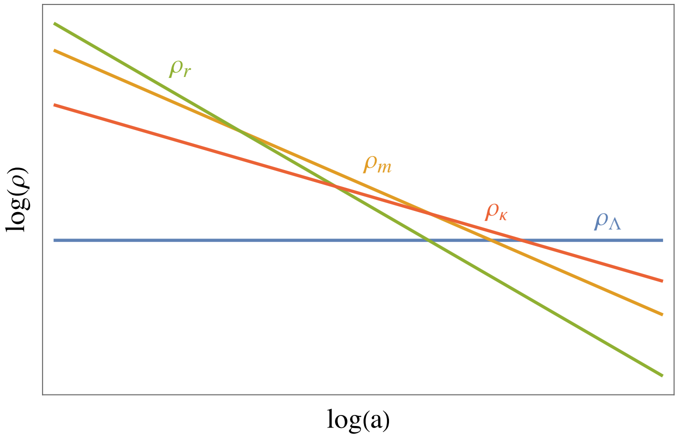
The equation of state $w = -1/3$ is also consistent with the continuity equation $\rho_\kappa \sim a^{-3(1+w)} = a^{-2}$.

With these nice observations. We can rewrite the Friedmann and Raychaudhuri equations as

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\sum_i \rho_i \ , \\ 2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 &= -8\pi G \sum_i p_i \ , \end{aligned} \tag{85}$$

where the index i goes over possible species of fluid in our universe, including radiation ($i = r$), matter ($i = m$), vacuum ($i = \Lambda$) and curvature ($i = \kappa$).

With the species and their continuity equation (a dependence in ρ) introduced above, we can imagine the following picture. Radiation tends to make the most important contribution to the total energy density of universe at very small a. Then it could be taken over by matter. And then curvature. Finally, at very large a, the vacuum energy will be the most important. (Note: ρ_m and ρ_κ may not necessarily dominate the universe during intermediate a values if their abundance is too low.)



Subtracting the two equations in Eq. (85), we can get

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\sum_i(\rho_i + 3p_i) = -\frac{4\pi G}{3}\sum_i(1+3w_i)\rho_i \ . \tag{86}$$

This implies that if our universe is dominated by a species with $w > -1/3$, then its expansion is decelerating. The expansion of universe is decelerating if it is dominated by matter or radiation. The expansion of accelerating if the vacuum energy dominates.

Another way of deriving the continuity equation. Taking another linear combination of the two equations in Eq. (85) leads to

$$\frac{d}{dt}\left(\frac{\dot{a}}{a}\right) = \frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 = -4\pi G\sum_i(1+w_i)\rho_i \ . \tag{87}$$

On the other hand, we can also derive another expression for $\frac{d}{dt}\left(\frac{\dot{a}}{a}\right)$ by taking time derivative to the square root of Friedmann equation, which leads to

$$\frac{d}{dt}\left(\frac{\dot{a}}{a}\right) = \sqrt{\frac{8\pi G}{3}}\frac{1}{2\sqrt{\rho}}\frac{d\rho}{dt} = \frac{4\pi G}{3}\sum_i a\frac{d\rho_i}{da} \ . \tag{88}$$

In the last step, we have used the Friedmann equation again to take care of the $\sqrt{\rho}$ factor in the denominator.
Eqs. (87) and (88) can lead to the same continuity equation as Eq. (80).

11 Toy Models of the FRW Universe

Before talking about the real world, let's first do some warm up exercises by considering simplified scenarios where in each scenario there is single species of fluid in the universe. The exercise is to derive the time dependence in $a(t)$ and and Hubble parameter, as solution to the Friedmann equation.

- **Universe with only radiation (always expands).** For radiation we have $\rho_r(t) = \rho_r(t_0)a(t)^{-4}$, and the Friedmann equation takes the form

$$\frac{\dot{a}(t)}{a(t)} = \sqrt{\frac{8\pi G\rho_r(t)}{3}} = \sqrt{\frac{8\pi G\rho_r(t_0)}{3}}a(t)^{-2} \ . \tag{89}$$

After taking the square root, we keep the + sign which corresponds to an expanding universe, instead of a shrinking one. In the last step, everything under the square root are constant of time. If we set $t = t_0$ and using $a(t_0) = 1$, we get the Hubble parameter today (it does not tell the t dependence in H though, see below)

$$H_0 = \sqrt{\frac{8\pi G\rho_r(t_0)}{3}} \ . \tag{90}$$

Eq. (89) is a simple differential equation for a. With the boundary condition, $a(t_0) = 1$, the solution is

$$a(t) = \sqrt{1 + 2H_0(t - t_0)} \ . \tag{91}$$

We can supplement it with an initial condition of the expansion where $a(0) = 0$ – the universe is very small at the beginning of big bang. This fixes

$$2H_0t_0 = 1, \quad a(t) = \sqrt{2H_0t} \sim t^{1/2} \ . \tag{92}$$

We find the universe always expands. With this result, we can derive the Hubble parameter H at any time t,

$$H = \frac{\dot{a}}{a} = \frac{1}{2t} \ . \tag{93}$$

As time increases, the Hubble parameter, or expansion rate of the universe decreases. For a radiation dominated universe, the age of the universe can be found via $t = 1/2H$, once the Hubble parameter is known.

A more straightforward way of calculating the age of universe (can be any time in the history of universe, not just today) would be

$$t = \int_0^t dt' = \int_0^a \frac{da'}{\dot{a}'} = \int_0^{a(t)} \frac{da'}{a'H(a')} = \int_0^{a(t)} \frac{a'da'}{H_0} = \frac{a(t)^2}{2H_0} = \frac{1}{2H_0a^{-2}} = \frac{1}{2H} \ . \tag{94}$$

We can also compute the size of particle horizon, using Eq. (73),

$$d_H(t) = a(t)\int_0^t \frac{dt'}{a(t')} = a(t)\int_0^{a(t)} \frac{da'}{\dot{a}'a'} = a(t)\int_0^{a(t)} \frac{da'}{a'^2H} = a(t)\int_0^{a(t)} \frac{da'}{H_0} = \frac{a(t)^2}{H_0} \ . \tag{95}$$

In the third step, we have used Eq. (89). For today, with $t = t_0$, the size of particle horizon is $d_H(t_0) = 1/H_0$.

- **Universe with only matter (always expands).** We can repeat the above calculation for the universe with only matter. In this case, the continuity equation states $\rho_m(t) = \rho_m(t_0)a(t)^{-3}$, and the Friedmann equation takes the form

$$\frac{\dot{a}(t)}{a(t)} = \sqrt{\frac{8\pi G\rho_m(t)}{3}} = \sqrt{\frac{8\pi G\rho_m(t_0)}{3}}a(t)^{-3/2} = H_0a(t)^{-3/2} \ . \tag{96}$$

Again, with today's boundary condition, we find

$$a(t) = \left[1 + \frac{3}{2}H_0(t - t_0)\right]^{2/3} \ . \tag{97}$$

Supplementing this solution with the same initial condition $a(0) = 0$, we further find

$$a(t) = \left(\frac{3}{2}H_0 t\right)^{2/3} , \quad H_0 = \frac{2}{3t_0} . \tag{98}$$

We find the universe always expands. The resulting time dependence in the Hubble parameter is

$$H = \frac{\dot{a}}{a} = \frac{2}{3t} . \tag{99}$$

The resulting particle horizon is

$$d_H(t) = a(t)\int_0^t \frac{dt'}{a(t')} = a(t)\int_0^{a(t)} \frac{da'}{\dot{a}'a'} = a(t)\int_0^{a(t)} \frac{da'}{H_0\sqrt{a'}} = \frac{2a(t)^{3/2}}{H_0} . \tag{100}$$

Today, $d_H(t_0) = 2/H_0$.

- **"Empty" universe with only vacuum energy (always expands).** Next, we switch to the "empty" universe with only vacuum energy around. This case is a bit peculiar. Because the energy density of vacuum energy is a constant of time, the right-hand side of Friedmann equation is a constant, implying that the Hubble parameter is a constant throughout the evolution of time,

$$H_0 = \sqrt{\frac{8\pi G\rho_\Lambda}{3}} . \tag{101}$$

The Friedmann equation reads

$$\frac{\dot{a}(t)}{a(t)} = H_0 . \tag{102}$$

With today's boundary condition $a(t_0) = 1$, the solution is

$$a(t) = e^{H_0(t-t_0)} . \tag{103}$$

Like the previous two scenarios, the universe always expands. Unlike the previous two scenarios, the solution of $a(t)$ here is not compatible with the initial condition $a(0) = 0$. Instead, it gives $a(0) = e^{-H_0 t_0}$ which is nonzero unless $H_0 t_0 \to +\infty$. The way out is to define the beginning time of universe $t_i = -\infty$, so that $a(t_i) = 0$. Clearly, in this special case there is no relation between time and the Hubble parameter in this case. The Hubble remains a constant. As time evolves, the scale parameter grows exponentially (much faster than power law).

A peculiar feature of the empty universe is the absence of particle horizon. Indeed

$$d_H(t) = a(t)\int_{t_i}^t \frac{dt'}{a(t')} = a(t)\int_{-\infty}^t e^{-H_0(t'-t_0)}dt' = a(t)\frac{1}{H_0}(e^{-H_0(-\infty-t)} - 1) = +\infty . \tag{104}$$

There is sufficient time (in fact, infinite time) for light to travel from any point of the universe and reach us today.

Let's introduce another useful physical concept here, the **Event horizon**. It is defined as the distance between two observers beyond which they can never establish any causal contact in the future. In other words, it is the distance light can travel in the future with respect to time t

$$d_{EH}(t) = a(t) \int_t^{+\infty} \frac{dt'}{a(t')} = a(t) \int_t^{+\infty} e^{-H_0(t'-t_0)} dt' = \frac{1}{H_0} \ . \tag{105}$$

The result is t independent. This is the largest distance light can travel by giving it infinite amount of time in the future. As a result, any two observers separated by distances larger than $1/H_0$ will lose contact from each other forever. This happens if the universe is filled by vacuum energy.

Here is an intuitive understanding of the event horizon. Suppose at time t_1, two observer are separated by a distance L. Observer A wants to send light signal to the other (B). At an infinitesimal later time $t_1 + \delta t$, light has traveled a distance $c\delta t = \delta t$. At the same time, the observer B has become more further apart from A due to the exponential expansion of the universe, $\delta L = Le^{H_0 \delta t} - L > LH_0 \delta t$. Clearly, if $L > H_0^{-1}$, A and B would be more distant from each other at time $t + \delta t$ than at time t, because expansion wins over light traveling. In this case, A and B are outside the event horizon. The same argument works for any time t, thus the event horizon is a constant in this universe.

We can also check the above event horizon integral for radiation and matter dominated universe, and the results will be + infinity. That means in those universes, light still has enough time to travel across any distance in the future. As a result, the event horizon does not exist.

Particle horizon is about the past. Event horizon is about the future.

- **A universe that can recollapse.** As the last example, we consider a less simple example where the universe is filled with both matter and curvature. The Friedmann equation reads

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{c_m}{a^3} - \frac{c_\kappa}{a^2} \ , \tag{106}$$

where $c_m = 8\pi G \rho_m(t_0)/3$ and $c_\kappa = -8\pi G \rho_\kappa(t_0)/3 = \kappa/R^2$ are constants. Clearly, if the universe is flat $\kappa = 0$, then the right-hand side only has matter term. We have done this exercise above. If $\kappa = \pm 1$, the story is less trivial. However, in the limit $a \to 0$, the asymptotic form of the equation is the same as that for universe with only matter, thus yielding the same solution as Eq. (98) for very small a.

To solve this differential equation analytically, we introduce the **conformal time**,

$$\tau(t) = \int_0^t \frac{dt'}{a(t')}, \quad \Rightarrow \quad \frac{d\tau}{dt} = \frac{1}{a} \ , \tag{107}$$

which grows monotonically with t (a is always positive) and satisfies $\tau(0) = 0$, and define

$$a' \equiv \frac{da}{d\tau} = \frac{da}{dt}\frac{dt}{d\tau} = a\dot{a} \ . \tag{108}$$

Eq. (106) can be rewritten as

$$a'^2 = c_m a - c_\kappa a^2 \ . \tag{109}$$

a) In the case $\kappa = 1$ thus $c_\kappa > 0$, the solution to Eq. (109) is

$$a(\tau) = \frac{c_m}{c_\kappa} \sin^2\left(\frac{\sqrt{c_\kappa}}{2}\tau\right) \ , \tag{110}$$

where we have set the initial condition to be $a(0) = 0$. Clearly, the universe first expands until τ reaches $\pi/\sqrt{c_\kappa}$ after which the universe will start to collapse. The scale factor ends up to become zero again at $\tau = 2\pi/\sqrt{c_\kappa}$ when the universe crunches to a back point.

b) In the case $\kappa = -1$ thus $c_\kappa < 0$, the solution to Eq. (109) is

$$a(\tau) = \frac{c_m}{c_\kappa} \sinh^2\left(\frac{\sqrt{-c_\kappa}}{2}\tau\right) \ . \tag{111}$$

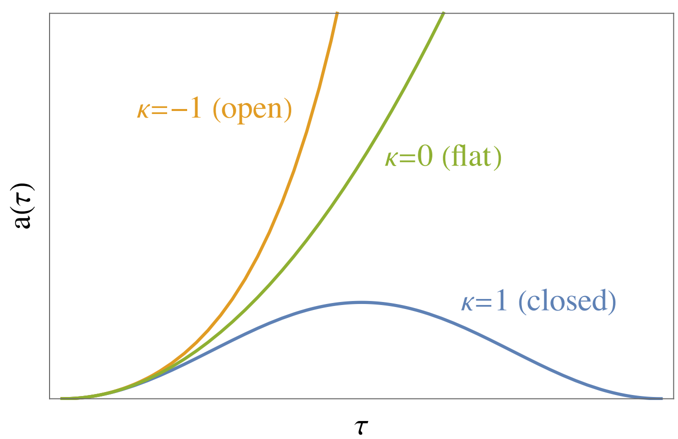
c) In the case $\kappa = 0$, we simply take over the solution in Eq. (98), but write it in term of the conformal time τ, where

$$\tau(t) = \int_0^t \frac{dt'}{a(t')} = \frac{2a(t)^{1/2}}{H_0} \ . \tag{112}$$

This gives

$$a(\tau) = \left(\frac{H_0\tau}{2}\right)^2 \ . \tag{113}$$

In the figure below, we plot $a(\tau)$ for the three cases, which correspond to closed, open and flat universe, respectively.



12 The ΛCDM Model

In this section, we present the working model that can successfully describe our universe – the ΛCDM model. In particular, it has a recipe for energy budget of each fluid in our universe. Let us first write the Friedmann equation Eq. (85) again,

$$\begin{aligned} H^2 = \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\left(\rho_r + \rho_m + \rho_\Lambda + \rho_\kappa\right) \\ &= \frac{8\pi G}{3}\left(\rho_{r,0}a^{-4} + \rho_{m,0}a^{-3} + \rho_{\Lambda,0} + \rho_{\kappa,0}a^{-2}\right) \ , \end{aligned} \tag{114}$$

where $\rho_{i,0}$ corresponds to the energy density of fluid i today. The Hubble parameter today is simply found by evaluating both sides at time t_0 with $a(t_0) = 1$,

$$H_0 = \sqrt{\frac{8\pi G}{3}\left(\rho_{r,0} + \rho_{m,0} + \rho_{\Lambda,0} + \rho_{\kappa,0}\right)} \ . \tag{115}$$

We introduce the critical density of our universe

$$\rho_c = \frac{3}{8\pi G} H_0^2 = 1.054 \times 10^{-5} h^2 \, \mathrm{GeV/cm}^3 \ , \tag{116}$$

where $h = 0.67$, and today's energy fraction for each fluid

$$\Omega_i = \frac{\rho_{i,0}}{\rho_c} \ , \tag{117}$$

where Ω_κ is the only one that can be negative. Clearly, $\Omega_r + \Omega_m + \Omega_\Lambda + \Omega_\kappa = 1$.

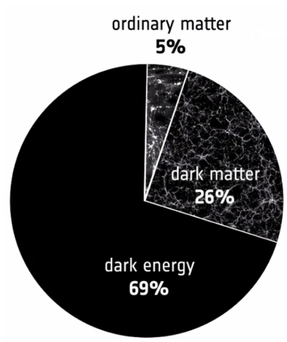
We can rewrite the Friedmann equation as

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \left(\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_\Lambda + \Omega_\kappa a^{-2}\right) \ . \tag{118}$$

In the equation, Ω_r, Ω_M, Ω_Λ, Ω_κ are the parameters describing our universe. I will present the experimental measured value of them, without explaining in detail how they are measured with the cosmological observation data (including CMB, BBN, large scale structure, etc). https://pdg.lbl.gov/2021/reviews/rpp2020-rev-astrophysical-constants.pdf

$$\Omega_\gamma \simeq 5.38 \times 10^{-5}, \quad \Omega_m = 0.315, \quad \Omega_\Lambda = 0.685, \quad \Omega_\kappa = (7 \pm 19) \times 10^{-4} \ . \tag{119}$$

Sometimes you see a figure like this.



"Cosmic Energy Budget"

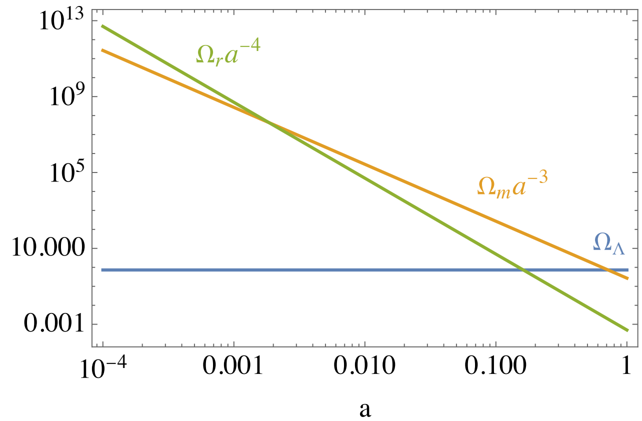
The fact that Ω_κ is very small implies that our universe is very close to being flat. From now on, we will simply set $\Omega_\kappa = 0$ in the discussions. The fact that Ω_γ for photon (not necessarily all the radiations today) is also very tiny means our universe today is mostly dominated by vacuum energy and matter. However, back in the early universe, radiation can become much more important, due to the most steep a dependence in ρ_r.

Another experiment input is the Hubble parameter today. We discussed how it could be measured back in Section 8. The measured value is

$$H_0 = 100h \, \mathrm{km/s/Mpc} = h \times (9.78 \, \mathrm{Gyr})^{-1} \ , \tag{120}$$

and $h = 0.674$.

Knowing the values of Ω_i and H_0 allow us to understand the past of our universe. Let's take a closer look at the a dependence in the energy densities. Similar to the $\rho_i(a)$ plot in Section 11, we now have the more realistic version as shown below.



The intersection point of different curves defines special times in the past. The matter-vacuum equality occurs when $\rho_m = \rho_\Lambda$. The corresponding scale factor can be found by

$$\Omega_m a^{-3} = \Omega_\Lambda, \quad \Rightarrow \quad a_{m\Lambda} = 0.78 \ . \tag{121}$$

Using the definition of redshift $1 + z = 1/a$, we find $z_{m\Lambda} = 0.28$.

The matter-radiation equality occurs when $\rho_m = \rho_r$. The corresponding scale factor can be found by

$$\Omega_m a^{-3} = \Omega_\gamma a^{-4}, \quad \Rightarrow \quad a_{eq} = 1.7 \times 10^{-4} \ . \tag{122}$$

Correspondingly, $z_{eq} = 5900$. This estimate is a bit off from reality because of the contribution from neutrinos as radiation in the past, which also contribute to ρ_r. As a result, the actual matter-radiation equality more recently, around $a_{eq} \simeq 3 \times 10^{-4}$ and $z_{eq} \simeq 3400$.

During the time window $a_{eq} < a < a_{m\Lambda}$, we have $\rho_m \gg \rho_r, \rho_\Lambda$, and the universe is matter dominated. In very early universe, $a < a_{eq}$, we have $\rho_r \gg \rho_m, \rho_\Lambda$, and the universe is radiation dominated. Only very recently, the vacuum energy takes over the dominance in the universe, but ρ_Λ has not become much larger than ρ_m yet.

With the above findings, we can finally show that our universe has been always expanding (assuming $\rho_\kappa \simeq 0$). From the Friedmann equation, we find $\dot{a} > 0$ throughout the radiation or matter dominated etas, as long as the universe was born to be expanding (the big bang). From Eq. (86), we can also find $\ddot{a} < 0$. Thus the expansion was decelerating during these eras but has not stopped at $a_{m\Lambda}$. Afterwards, then the universe becomes vacuum energy dominated, Eq. (86) tells $\ddot{a} > 0$. The expansion of universe becomes accelerating. Putting everything together, the universe was always expanding.

Now we are ready to calculate some important quantities characterizing our universe.

- **Age of the Universe.** We can evaluate the age at any time t, including present, past and future. Using Eq. (118),

$$t = \int_0^t dt' = \int_0^a \frac{da'}{\dot{a}'} = \int_0^a \frac{da'}{a' H_0 \sqrt{\sum_i \Omega_i a'^{-3(1+w_i)}}} = \int_0^a \frac{da'}{a' H_0 \sqrt{\Omega_\Lambda + \Omega_m a'^{-3} + \Omega_r a'^{-4}}}$$
$$= \frac{1}{H_0} \int_0^a \frac{da'}{\sqrt{\Omega_\Lambda a'^2 + \frac{\Omega_m}{a'} + \frac{\Omega_r}{a'^2}}} \ . \tag{123}$$

The age of universe today corresponds to $a = a_0 = 1$. With the input values Eq. (119), the integral can be completed, leading to

$$t_0 = \frac{0.95}{H_0} = 13.8\,\text{Gyr} \ , \tag{124}$$

where Gyr $= 10^9$ year.

We can also evaluate the two special times in early universe, corresponding to matter-vacuum and matter-radiation equality. Applying Eqs. (121) and (122) in the upper limit of the integral, they are

$$t_{m\Lambda} = \frac{0.72}{H_0} = 10.4\,\text{Gyr} \ , \quad t_{eq} = \frac{1.54 \times 10^{-6}}{H_0} = 2.2 \times 10^4\,\text{yr} \ . \tag{125}$$

Clearly, matter-radiation equality occurs when our universe was very "young".

- **The particle horizon.** We calculate the size of particle horizon using Eq. (73),

$$d_H(t) = a(t) \int_0^t \frac{dt'}{a(t')} = a(t) \int_0^{a(t)} \frac{da'}{a' \dot{a}'} = \frac{a}{H_0} \int_0^a \frac{da'}{\sqrt{\Omega_\Lambda a'^4 + \Omega_m a' + \Omega_r}} \ . \tag{126}$$

The size of particle horizon today is (use $a(t_0) = 1$)

$$d_H(t) = \frac{1}{H_0} \int_0^1 \frac{da'}{\sqrt{\Omega_\Lambda a'^4 + \Omega_m a' + \Omega_r}} = \frac{3.19}{H_0} = 14\,\text{Gpc} = 46 \text{ billion light-years} \ . \tag{127}$$

Currently, and in the past, it grows with expansion of the universe faster than linear in a. In Section 11, we have shown that for radiation (matter) dominated universe $d_H \propto a^2$ ($a^{3/2}$). Thus more and more stuff are coming into our particle horizon, which means the light they emitted in past could become visible to us. In the far future, $t \gg t_0$ (thus $a \gg 1$), when the universe is dark energy dominated, the asymptotic behavior will be $d_H \propto a$. Eventually, no more stuff will enter the particle horizon.

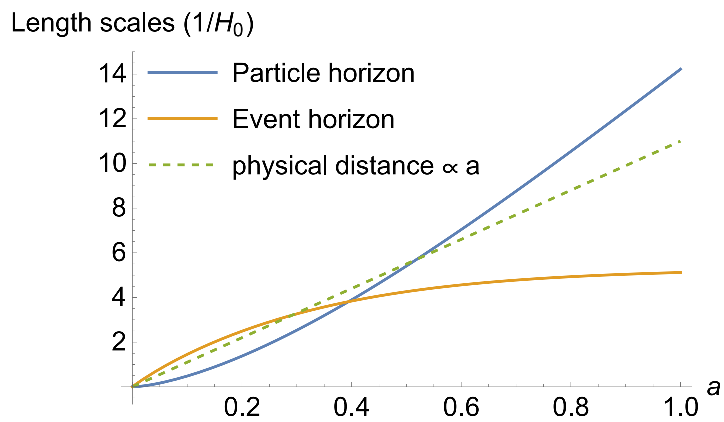
- **The event horizon.** Thanks to the presence of dark energy, our universe also has an event horizon.

$$d_{EH}(t) = a(t) \int_t^{+\infty} \frac{dt'}{a(t')} = a \int_a^{+\infty} \frac{da'}{a' \dot{a}'} = \frac{a}{H_0} \int_a^{+\infty} \frac{da'}{\sqrt{\Omega_\Lambda a'^4 + \Omega_m a' + \Omega_r}} \tag{128}$$

The value today is

$$d_{EH}(t_0) = \frac{1}{H_0} \int_1^{+\infty} \frac{da'}{\sqrt{\Omega_\Lambda a'^4 + \Omega_m a' + \Omega_r}} = \frac{1.15}{H_0} = 5.1\,\text{Gpc} \ . \tag{129}$$

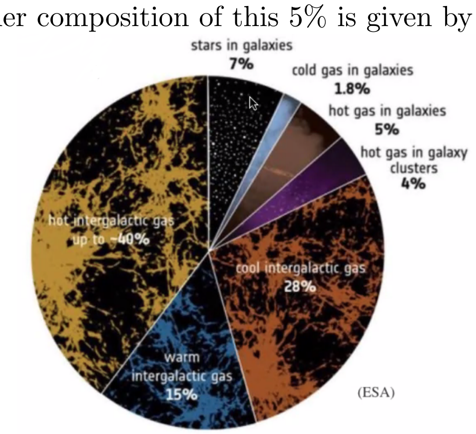
Remarkably, the event horizon always grows at a rate lower than a. In the far future, d_{EH} will approach to a constant value. This means more and more stuff are constantly moving beyond our event horizon. The light they emit after they leave the event horizon will never reach to us (although we could still see their light from the past if they lie within the particle horizon).



13 Thermodynamics of Expanding Universe

So far, we only talked about the dynamics of the expansion of our universe. The detailed behavior of expansion depends on which type of fluid dominates the energy of the universe. The impact of each fluid on the expansion rate is purely gravitational.

In the rest of this course, we will talk about other interactions (electromagnetic, strong, and weak) beyond gravity. These are the interactions for elementary particles. To our knowledge, at least part of the stuff in the universe are made of known elementary particles, including atoms, photons, and neutrinos. Today, they make up about 5% of the energy budget (see figure below Eq. (119)). They are sometimes referred to as the **visible sector** of the universe. The further composition of this 5% is given by the following figure.



Stars, galaxies and gases are all made of atoms. At more fundamental level, they are made of nuclei (protons, neutrons) and electrons. In addition, the CMB photons are also particles. Back in the early universe, these particles interact with each other and establish thermal equilibrium, **with a temperature**. In this case, the distribution of a large number of these particles are governed by the law of thermodynamics.

The phase space distribution function $f(\vec{x}, \vec{q}, t)$ describes the number of particles in a particular state (having momentum $\vec{q}$ and located at position $\vec{x}$). In general there are 7 variables. For homogeneous and isotropic universe, the $\vec{x}$ dependence drops out. Moreover f could only depend on the magnitude but not the direction of $\vec{q}$. Using the on-shell relation, $E = \sqrt{|\vec{q}|^2 + m^2}$, the phase space distribution of a particle in the expanding universe can be written as

$$f(E,t) \ . \tag{130}$$

We can calculate a number of quantities with a phase space distribution.

Number density:

$$n(t) = g \int \frac{d^3\vec{q}}{(2\pi)^3} f(E,T) \ , \tag{131}$$

where the factor g counts the number of degrees of freedom in the particle (e.g. spin). In an expanding universe, the time dependence is encoded in $T(t)$ (more discussions on this are coming shortly).

Energy density:

$$\rho(t) = g \int \frac{d^3\vec{q}}{(2\pi)^3} E f(E,T) \ . \tag{132}$$

Pressure:

$$p(t) = g \int \frac{d^3\vec{q}}{(2\pi)^3} \frac{|\vec{q}|^2}{3E} f(E,T) \ . \tag{133}$$

We can make further progress with the integrals if a particle is part of a thermal bath (made of a large number of similar particles) with temperature T. In this case, the only two possible phase space distribution function are

$$f^{\rm eq}(E,T) = \frac{1}{e^{(E-\mu)/T} \pm 1} \ , \tag{134}$$

where $+$ sign is for fermion and $-$ sign for boson, and μ is called chemical potential. These are called Fermi-Dirac and Bose-Einstein distribution functions.

For relativistic particles ($m \ll T$) with zero chemical potential, we have

$$n^{\rm eq} = \begin{cases} g\zeta(3)T^3/\pi^2 & \text{boson} \\ (3/4)g\zeta(3)T^3/\pi^2 & \text{fermion} \end{cases} \tag{135}$$

and

$$\rho^{\rm eq} = \begin{cases} \pi^2 g T^4/30 & \text{boson} \\ (7/8)\pi^2 g T^4/30 & \text{fermion} \end{cases} \tag{136}$$

and

$$p^{\rm eq} = \rho^{\rm eq}/3 \ . \tag{137}$$

This we can conclude that **relativistic particles make up a radiation fluid.**

On the other hand, for non-relativistic particles ($m \gg T$), the thermal number density becomes

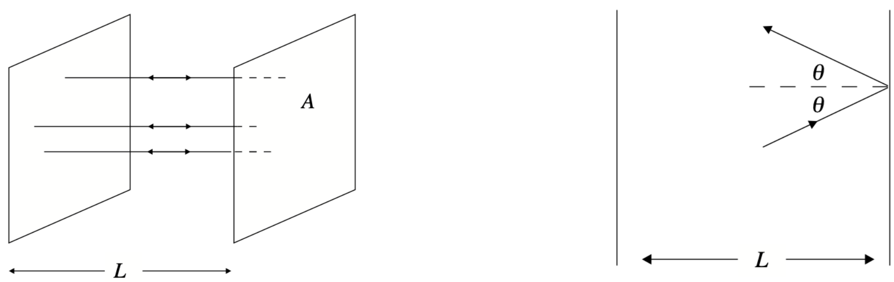
$$n^{\rm eq} \simeq g \left(\frac{mT}{2\pi}\right)^{3/2} e^{-(m-\mu)/T} \ , \tag{138}$$

which is identical for boson and fermion. The number density is strongly suppressed by the exponential factor. (The exponential suppression eventual saturates once the heavy particle drops out of thermal equilibrium, which is called freeze out. If it decays, one needs to add another exponential factor.) The corresponding energy density and pressure are

$$\rho^{\rm eq} \simeq \left(m + \frac{3}{2}T\right) n^{\rm eq}, \quad p^{\rm eq} \simeq T n^{\rm eq} \ll \rho^{\rm eq} \ . \tag{139}$$

The pressure is highly suppressed if $T \ll m$. Therefore, **non-relativistic particles make up matter fluid.**

A derivation of the pressure formula.
Pressure is force per area, or momentum transfer per area per time. Look at the picture below (left), where a particle with momentum $\vec{q}$ scatters inside a box:



Each time the particle hits the wall and bounces back, the change in momentum is $\Delta q = 2|\vec{q}|$. The time it takes for the particle to have another bounce (on the same wall) is the time it travels from one wall to the other wall and back, $\Delta t = 2L/v$. Therefore, the pressure exerted by this particle is

$$p = \frac{\Delta q}{\Delta t A} = \frac{|\vec{q}|v}{LA} = \frac{|\vec{q}|v}{V} \ , \tag{140}$$

where V is the volume. This result is actually wrong because the scattering with wall is not always head-on, but could happen with an angle θ. See the right plot above. In this case, we should modify $\Delta q = 2|\vec{q}|\cos\theta$ and $\Delta t = 2L/(v\cos\theta)$. As a result,

$$p = \frac{|\vec{q}|v\cos^2\theta}{V} \ . \tag{141}$$

Averaging over all possible angles, $\int_0^\pi \sin\theta d\theta \cos^2\theta = 1/3$. Therefore, the realistic pressure in three dimension is

$$p = \frac{|\vec{q}|v}{3V} \ . \tag{142}$$

Now, imagine there is more than one particle around, with a phase space distribution $f(E)$, where $E = \sqrt{|\vec{q}|^2 + m^2}$. Inside volume V, there are $N = V\int \frac{d^3\vec{q}}{(2\pi)^3} f(E)$. We need to add up all their contributions to the pressure,

$$p = V\int \frac{d^3\vec{q}}{(2\pi)^3} f(E) \frac{qv}{3V} = \int \frac{d^3\vec{q}}{(2\pi)^3} f(E) \frac{|\vec{q}|^2}{3E} \ . \tag{143}$$

In the last step we used the relation $v = |\vec{q}|/E$. This proves Eq. (133).

Let's now imagine several different kinds of particles are in thermal equilibrium with each other in the universe. (Note: there is a large number of each kind of particle around.)

They share the same temperature T. The first thing cosmology cares is the total energy density, which appears on the right-hand side of the Friedmann equation. Clearly, relativistic particles make more important contribution, because the number density for non-relativistic particle is exponentially suppressed. The total energy density can be written as

$$\begin{aligned} \rho_{\rm tot} &= \frac{\pi^2}{30} g_* T^4 \ , \\ g_* &= \sum_{i={\rm bosons}} g_i + \frac{7}{8} \sum_{i={\rm fermions}} g_i \ . \end{aligned} \tag{144}$$

One should note that g_* is a function of temperature. The sums only count light particles with $m \ll T$. For another temperature, some particle could become non-relativistic, and vice versa. As a result g_* could change.

More generally, the universe could be radiation dominated with several species each having temperature T_i. In this case, g_* is

$$g_* = \sum_{i={\rm bosons}} \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{i={\rm fermions}} g_i \left(\frac{T_i}{T}\right)^4 \ . \tag{145}$$

Hereafter, when mentioning T, I always mean the temperature of photons.

The Friedmann equation, $H^2 = 8\pi G \rho_{\rm tot}/3$, then tells us the Hubble parameter,

$$H \simeq \frac{1.66\sqrt{g_*} T^2}{M_{\rm pl}} \ , \tag{146}$$

where $M_{\rm pl} = 1/\sqrt{G} = 1.2 \times 10^{19}\,{\rm GeV}$. The Hubble parameter is also called the **expansion rate** of the universe.

What also drive the evolution of universe are other rates, especially the **reaction rates** for elementary particles. Imagine a microscopic process where two particles A and B can interact and turn into other states. When these two particles are put together (they can do so by traveling to each other in the universe), how often they react is described by a quantity called cross section σ. σ can be fully determined once the fundamental physics behind the interaction is known, e.g. electromagnetic interaction, weak interaction. For each A particles, it sees a number density of particle B around, denoted by n. The reaction rate per A particle is then calculated using

$$\Gamma = n\sigma v \ , \tag{147}$$

where v is the relative velocity A and B are traveling toward each other.

When there are two rates H and Γ, we compare them with each other.

If $\Gamma \gg H$, the particle reaction is likely to occur many times within the Hubble time H^{-1}. (Remember: rate $\times$ time is the number of reactions that can occur.) In this case, the reaction is fast, and all the particles involved in the initial or final states (e.g. A) will be in thermal equilibrium with each other, and can have a common temperature T.

If $\Gamma \ll H$, the reaction ceases to occur. Note it does not shuts off strictly unless $\Gamma = 0$. What it means is there are a large number of particle A around and only a small fraction of them would have a chance react. In this case, A exists from the thermal equilibrium.

We will see such comparisons over and over again when discussing recombination, neutrino decoupling, and big-bang nucleosynthesis sections. It will play a key role in determining what the universe looks.

Entropy conservation.

There is a conserved quantity during the expansion of the universe – the total entropy. Taking any comoving volume in the universe, there is no heat transfer into or out of this volume during its expansion. (The universe is homogeneous.) In other words, the universe expands adiabatically. This implies, $dQ = TdS = 0$, where S can be the entropy inside any comoving volume.

The first law of thermodynamics states

$$dE = TdS - pdV + \sum_i \mu_i dN_i \ , \tag{148}$$

where S is the entropy of the system, and i labels particle species. Let's next take $E = E(T, V)$, $S = S(T, V)$ and $N = N(T, V)$ with T and V as independent variables. This allow us to expand Eq. (148) as

$$\frac{\partial E}{\partial T}dT + \frac{\partial E}{\partial V}dV = T\left(\frac{\partial S}{\partial T}dT + \frac{\partial S}{\partial V}dV\right) - pdV + \sum_i \mu_i \left(\frac{\partial N_i}{\partial T}dT + \frac{\partial N_i}{\partial V}dV\right) \ . \tag{149}$$

In the equation, the coefficients of dT and dV must vanish separately,

$$\begin{aligned} \frac{\partial E}{\partial T} &= T\frac{\partial S}{\partial T} + \sum_i \mu_i \frac{\partial N_i}{\partial T} \ , \\ \frac{\partial E}{\partial V} &= T\frac{\partial S}{\partial V} - p + \sum_i \mu_i \frac{\partial N_i}{\partial V} \ . \end{aligned} \tag{150}$$

The second equation is most useful here. Applying this to the universe by choosing any volume V, and define the densities (all of them are extensive quantities, proportional to volume)

$$\rho = \frac{E}{V} = \frac{\partial E}{\partial V}, \quad s = \frac{S}{V} = \frac{\partial S}{\partial V}, \quad n_i = \frac{N_i}{V} = \frac{\partial N_i}{\partial V} \ , \tag{151}$$

lead to

$$\rho + p - Ts - \sum_i \mu_i n_i = 0 \ . \tag{152}$$

This gives solution to the entropy density s

$$s = \frac{\rho + p - \sum_i \mu_i n_i}{T} \ . \tag{153}$$

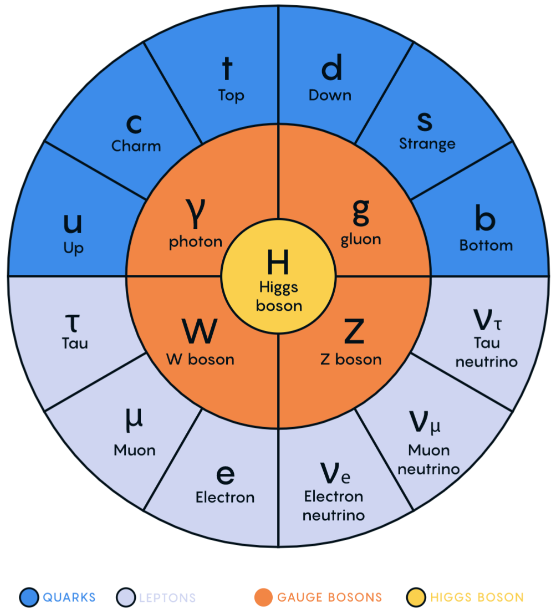
Throughout the history of universe, the total entropy, or the quantity sa^3, is conserved.

In a radiation dominated universe with several species each having temperature T_i, we have (assuming $\mu_i = 0$)

$$\begin{aligned} s &= \frac{2\pi^2}{45} g_{*S} T^3 \ , \\ g_{*S} &= \sum_{i=\text{bosons}} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8} \sum_{i=\text{fermions}} g_i \left(\frac{T_i}{T}\right)^3 \ . \end{aligned} \tag{154}$$

Temperature of known particles.

All Standard Model particles are in thermal equilibrium via gauge interactions (strong, weak, electromagnetic) at very high temperatures. All T_i are equal to the photon temperature T, when $T \gg$ MeV.



Temperature evolution in early universe.

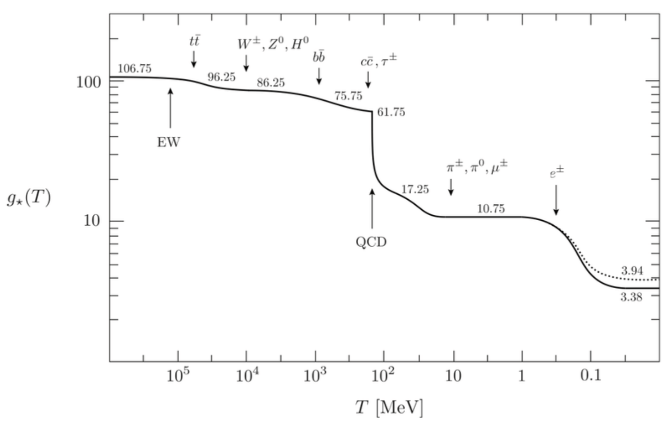
From the conservation of entropy, we can derive the evolution of temperature. Consider a comoving volume $V = a^3$. The total entropy inside this volume is

$$S = \frac{2\pi^2}{45} g_{*S} T^3 a^3 \ . \tag{155}$$

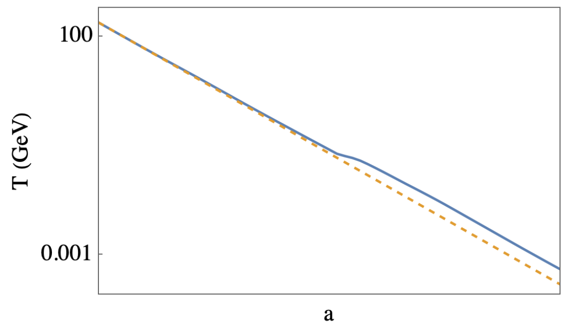
We know S is conserved. Therefore, if g_{*S} is a constant, the temperature falls inversely with the expansion of universe,

$$T \sim \frac{1}{a} \ . \tag{156}$$

For temperature between electroweak scale (100 GeV) and MeV scale, g_{*S} does change with T most drastically when the temperature crosses each mass of Standard Model particles. See plot below.



As a result, the temperature falls a bit slower than $1/a$. See plot below. When a particle becomes heavy, its population gets exponentially suppressed. These particles annihilate or decay into lighter particles, resulting in the latter becoming relatively hotter than the $T \sim 1/a$ cooling case (Note T still decreases monotonically with the expansion).



Note on photon temperature after decoupling.
Before photon-electron decoupling (when recombination occurs), photons are in thermal equilibrium and their population follows the Bose-Einstein distribution

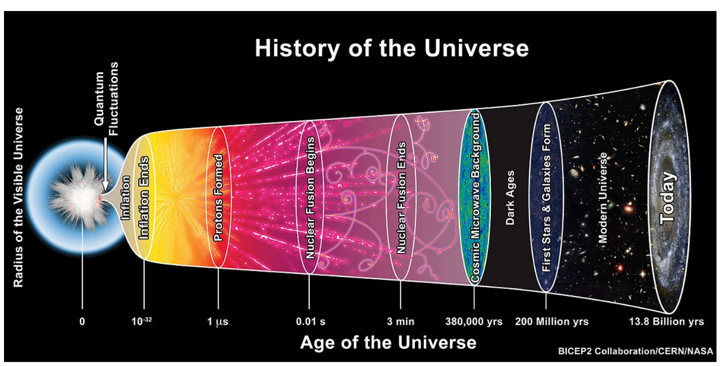
$$f(E,T) = \frac{2}{e^{E/T}-1} \ . \tag{157}$$

After the decoupling, the photons are no longer in thermal equilibrium with any thing else. They do not self interact either. Question: can we continue to use the above Bose-Einstein distribution to describe photon? The answer is yes.

In Section 8, we have discussed that as the universe expands, the wavelength of photon grows with a. In turn, its energy redshifts (decreases) as $E \sim 1/a$. In the previous section, we used entropy conservation to show that the temperature of photon decreases as $T \sim 1/a$. As a result, the ratio E/T remains a constant of time, and the form of Bose-Einstein distribution remains to be valid for late times.

This is why we still use Eq. (157) when talking about CMB photons today and say it has a temperature ($T = 2.7\,\mathrm{K}$).

14 Chronology of the universe



Because the photon temperature T falls monotonically with the expansion of the universe, sometimes the falling T is used to indicate the arrow of time. Here is a list of the most important events in the history of universe.

- The big-bang occurs, beginning of time.
- There is likely a stage of inflation and reheating.

- At very high temperatures $T \gg 100\,\text{GeV}$, all Standard Model particles are relativistic and in thermal equilibrium with each other.
- By the time $T = 1\,\text{GeV}$ ($t = 1\,\mu\text{s}$), quarks start to form protons and neutrons. Many Standard Model particles already became heavy and non-relativistic.
- At $T = 10\,\text{MeV}$, the only relativistic particles in the universe are photon, electron and neutrinos.
- Around $T = 1\,\text{MeV}$ ($t = 1\,\text{s}$), neutrinos decouple from weak interaction with electrons.
- In the window $10\,\text{keV} < T < 1\,\text{MeV}$, the big-hang nucleosynthesis (BBN) takes place allowing protons and neutrino to form nuclei.
- Around $T = 1\,\text{eV}$, matter-radiation equality. The large-scale structure (LSS) starts to grow efficiently. Dark matter is needed for structure formation.
- Around $T = 0.3\,\text{eV}$ ($t = 370,000\,\text{years}$), recombination takes place allowing protons and electrons to form the hydrogen atom. Photon decouples from electrons. The universe becomes transparent to photons afterward. The free propagating photons form the cosmic microwave background (CMB).
- Around $T = 60\,\text{K}$ ($t = 10^8 - 10^{10}\,\text{years}$), galaxies and stars form. There is a stage of re-ionization which I will not talk too much about.
- Around $T = 3.5\,\text{K}$, matter-dark-energy equality. The universe starts accelerating expansion.
- Temperature of photon continue to cool until today $T = 2.7\,\text{K}$.

15 Relic Neutrinos

So far, we have collected sufficient tools in cosmology (including how the universe expands and how to describe particles in it). Time comes to discuss several important epochs in the early universe. Let's first dial the clock back to the time when the universe is slightly younger than 1 second. At this time, the universe is radiation dominated with a photon temperature of around few MeV. The most populated particles in the universe include photon, electron/positron, neutrinos. They are all radiation species at this time because $T \gg m$. All of them are still in thermal equilibrium.

In particular, neutrinos and electrons are in thermal equilibrium via weak interactions, via processes like $\nu\bar{\nu} \leftrightarrow e^+e^-$. The weak interaction cross section can be estimated as $\sigma \sim G_F^2 E_\nu^2 \simeq G_F^2 T^2$, where G_F is the Fermi's constant for weak interaction. In the last step, we approximated the neutrino energy with its temperature. Now imagine you are a neutrino. You look around, there is a number density of antineutrinos around you. They are relativistic. Thus $n_{\bar{\nu}} \sim T^3$. All of them could have weak interaction with you, and how likely the reaction occurs is dictated by the above cross section. The weak interaction per neutrino is then

$$\Gamma = n\sigma v \sim G_F^2 T^5 \ . \tag{158}$$

where $v \simeq c$ for relativistic neutrinos. Meanwhile, the expansion rate of the universe is

$$H \simeq \frac{T^2}{M_{\text{pl}}} \ . \tag{159}$$

Comparing the two, we concludes that Γ decreases with T (or the expansion of universe) faster than H. Thus the weak interaction will *decouple* at temperature when $\Gamma = H$, which leads to

$$T_d \sim \left(\frac{1}{G_F^2 M_{\rm pl}}\right)^{1/3} \sim 0.8\,{\rm MeV}\ . \tag{160}$$

where we used $G_F = 1.16 \times 10^{-5}\,{\rm GeV}^{-2}$, $M_{\rm pl} = 1.2 \times 10^{19}\,{\rm GeV}$. At later time $T < T_d$, we will have $\Gamma < H$, thus there is no time for weak interaction to occur any more. As an approximation, we simply assume no weak interaction happens after T_d any more. (This is not strictly true because probability being smaller than one does not mean the process cannot happen, especially given the presence of larger number of neutrinos in the universe. The more precise of calculating neutrino decoupling is to solve the Boltzmann equation. We will not resort to that intricacy here in this course.)

Decoupling implies that the neutrino fluid will stop having heat transfer with the electron-photon fluid at lower temperature. (Electron and photon are in thermal equilibrium via electromagnetic interaction which is much stronger and does not decouple until much later.) They decouple from each other. Although they still have the same temperature and both scale as $1/a$ with the expansion of universe, right after decoupling. The two temperatures no longer need to be identical if some physics happens at later time.

Next, let time elapse a bit further until the universe cools to photon temperature equal to the electron mass, equal to 511 keV (or roughly half an MeV). Until now, we still have $T_\gamma = T_\nu$ because nothing has happened yet. As T_γ drops well below the electron mass, the number density of electrons will become Boltzmann suppressed if they stay in thermal equilibrium and they do. This is called electron decoupling. At fundamental level, it is the annihilation process $e^+e^- \to \gamma\gamma$ that occurs in the forward direction to deplete the electron number density. (At higher temperatures, the inverse process was equally important to keep the electron number density in tact, but not any more when electron mass becomes larger than the temperature (or $k_B T_\gamma$).)

It is important to note that e^+e^- will only annihilate into photons but not neutrinos because the latter has already decoupled. Entropy conservation for the electron-photon fluid leads to the equation

$$\frac{2\pi^2}{45}\left(2 + \frac{7}{8} \times 2 \times 2\right) T_d^3 a(T_d)^3 = \frac{2\pi^2}{45} \times 2 \times T_\gamma^3 a(T_\gamma)^3\ , \tag{161}$$

where $T_d = 1\,{\rm MeV}$ stands for a time before electron decoupling, whereas we choose $T_\gamma \ll m_e$ stands for a time after electron decoupling.

On the other hand, nothing happens to the neutrino fluid, and we have

$$\frac{2\pi^2}{45}\left(\frac{7}{8} \times 3 \times 2\right) T_d^3 a(T_d)^3 = \frac{2\pi^2}{45}\left(\frac{7}{8} \times 3 \times 2\right) T_\nu^3 a(T_\gamma)^3\ , \tag{162}$$

where both neutrino and electron have the same temperature right after weak interaction decoupling when the universe has a scale factor $a(T_d)$. And the later time T_ν is evaluated at the same time as T_γ when the universe has scale factor $a(T_\gamma)$.

Dividing the above two equations we get

$$\frac{T_\gamma}{T_\nu} = \left(\frac{11}{4}\right)^{1/3} \simeq 1.4\ . \tag{163}$$

In standard cosmology, this temperature ratio always hold after the electron decoupling.

16 Big-Bang Nucleosynthesis

Now let's still consider the universe around 1 second old.

In addition to photons and neutrinos, there are also some smaller population of proton and neutron (they are called baryons). (Neutrons are unstable but have a lifetime of 10 minutes, so there is no time for them to decay yet.) Naively, one would expect the presence of these baryons to be totally negligible because they have mass around GeV scale, thus the thermal population would be suppressed by the Boltzmann factor $e^{-M/T} \sim e^{-1000} \approx 0$. However, our universe features a baryon asymmetry such that the number of proton/neutron and their anti-particles are not equal. Such an asymmetric number density does not get Boltzmann suppressed but simply dilutes with the expansion of universe as a^{-3}. This behavior is the same as the number density of photons after the electron decoupling. Therefore, we have a conserved quantity

$$\eta_b = \frac{n_b}{n_\gamma} \simeq 6 \times 10^{-10} \ . \tag{164}$$

There are no antibaryons (anti atoms) around, $n_{\bar{b}} = 0$. I will not explain how the baryon asymmetry was created, but simply use this number as the initial condition for big-bang nucleosynthesis (BBN).

Chemical potential for non-relativistic particle in thermal equilibrium.
For a non-relativistic particle in thermal equilibrium, and having a particle-anti-particle asymmetry in the number density, Eq. (138) states the number density should take the form

$$n^{\rm eq} \simeq g \left(\frac{mT}{2\pi}\right)^{3/2} e^{-(m-\mu(T))/T} \ . \tag{165}$$

Here the chemical potential must be temperature dependent, and takes the proper form such that $n^{\rm eq} \sim a^{-3}$. Applying this to the case of baryon asymmetry, where

$$n_b^{\rm eq} = \eta_b n_\gamma = \eta_b \frac{2\zeta(3)T^3}{\pi^2} \ . \tag{166}$$

This allows us to solve for $\mu(T)$,

$$\mu(T) = m + T \log\left[\eta_b \frac{2^{3/2}\zeta(3)}{\sqrt{\pi}} \left(\frac{T}{m}\right)^3\right] \ . \tag{167}$$

BBN is the process where strong interaction plays an important role and allow protons and neutrons to bind into nuclei. The corresponding nuclear fusion processes happen over a period when the age of universe is between 1 second and $\sim$ 100 minutes. Note $H = 1/(2t) \sim T^2$. The corresponding temperature range is between 1 MeV and 10 keV. Clearly, while the nuclear fusion processes take place, the universe expands and electron decouples in the middle of it, and more over, there is long enough time for neutron to decay. There is a close interplay and rich physics among all these effects.

Here we give a simplified description of the timeline and essential physics.

First, before nuclear fusion occurs, weak interaction has already decoupled. Before the decoupling, proton and neutron are in thermal equilibrium through the process

$$n + \nu_e \leftrightarrow p + e^- \ . \tag{168}$$

Using Eq. (138), we derive that immediately after weak interaction decoupling

$$
\begin{aligned}
n_p^{\mathrm{eq}} &= g_p \left(\frac{m_p T}{2\pi}\right)^{3/2} e^{-(m_p-\mu_p)/T} \ , \\
n_n^{\mathrm{eq}} &= g_n \left(\frac{m_n T}{2\pi}\right)^{3/2} e^{-(m_n-\mu_n)/T} \ ,
\end{aligned}
\tag{169}
$$

where $g_p = g_n = 2$ counts the spin degrees of freedom for proton and neutrino, respectively. Using the chemical equilibrium relation

$$
\mu_n + \mu_{\nu_e} = \mu_p + \mu_e \ , \tag{170}
$$

and observing that $\mu_e \approx \mu_\nu \approx 0$ (Note neutrino has just decoupled from weak interaction, and electron is still in thermal equilibrium; Both follow the regular Fermi-Dirac distribution without chemical potential.),

$$
\left.\frac{n_n}{n_p}\right|_{T=T_d} = \left(\frac{m_n}{m_p}\right)^{3/2} e^{-(m_n-m_p)/T_d} \simeq \frac{1}{5} \ . \tag{171}
$$

where we used $m_n = 939.6\,\mathrm{MeV}$, $m_p = 938.3\,\mathrm{MeV}$, and $T_d = 0.8\,\mathrm{MeV}$.

Second, to approach the problem of nucleosynthesis here we want to make simplifications. This is mainly because there is a whole periodic table of nuclei that are made of protons and neutrons. Although not all of them are formed during BBN; only the smaller ones are formed up to Lithium, there are still too many. So let's first oversimplify problem by considering only the smallest nucleus, the deuteron (or deuterium), which is made of proton and neutron bounded together by strong force. You could think of the analogy of hydrogen atom. And yes, there is a binding energy for deuteron, $\Delta \simeq 2.2\,\mathrm{MeV}$. It is then important to point out that deuteron is only efficiently formed when the temperature of photons in the universe falls well below this binding energy (roughly, an order of magnitude). Otherwise, an energetic photon from the thermal plasma can strike on a deuteron and dissociate it into free proton and neutron. Thus one needs to be patient and wait for the temperature to drop due to the expansion of the universe.

However, we should also keep in mind that neutron's lifetime is about 10 minutes, thus there is not lots of time to wait otherwise there would be no neutrons around to fuel the deuteron. Because deuterons are made of equal number of protons and neutrons, Eq. (171) tells that at most 1/5 of the primordial protons can end up in deuteron. I.e., $n_D/n_p < 1/4$.

Let us now be more quantitative about the above discussions. The key microscopic process for forming the deuteron D is

$$
p + n \to D + \gamma \ . \tag{172}
$$

Let's assume this process is thermal equilibrium so everybody involved in the process shares the same temperature as the photon.

Using the in-equilibrium number density Eq. (138) again,

$$
\begin{aligned}
n_p^{\mathrm{eq}} &= g_p \left(\frac{m_p T}{2\pi}\right)^{3/2} e^{-(m_p-\mu_p)/T} \ , \\
n_n^{\mathrm{eq}} &= g_n \left(\frac{m_n T}{2\pi}\right)^{3/2} e^{-(m_n-\mu_n)/T} \ , \\
n_D^{\mathrm{eq}} &= g_D \left(\frac{m_D T}{2\pi}\right)^{3/2} e^{-(m_D-\mu_D)/T} \ ,
\end{aligned}
\tag{173}
$$

and the mass relation (Introduce the binding energy Δ. From the plot above, $\Delta \simeq 2.2\,\text{MeV}$ for deuteron)

$$m_p + m_n = m_D + \Delta \ , \tag{174}$$

and the chemical equilibrium condition (photon has no chemical equilibrium)

$$\mu_p + \nu_n = \mu_D \ , \tag{175}$$

we can derive

$$\frac{n_D}{n_p n_n} = \frac{g_D}{g_p g_n} \left(\frac{m_D}{m_p m_n} \right)^{3/2} \left(\frac{T}{2\pi} \right)^{-3/2} e^{\Delta/T} \ . \tag{176}$$

Proton and neutron are spin 1/2 thus $g_p = g_n = 1/2$. Deuteron has spin 1 and is massive, thus $g_D = 3$. Using the approximate mass relation $m_D \simeq m_p + m_n$, we obtain the **Saha equation**

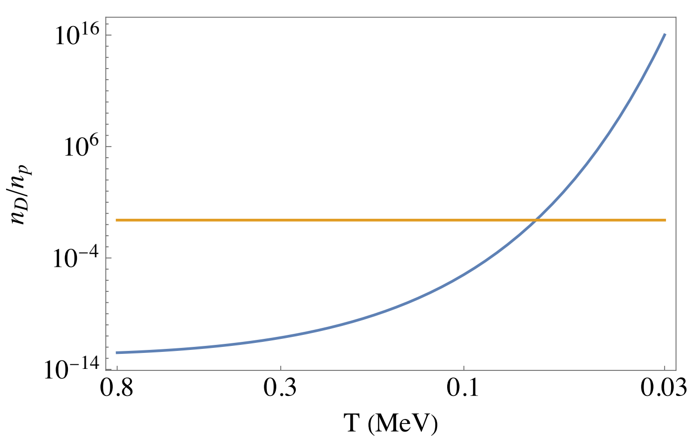
$$\frac{n_D}{n_p n_n} = 6 \left(\frac{\pi}{m_p T} \right)^{3/2} e^{\Delta/T} \ . \tag{177}$$

This equation is valid during BBN when the nuclear fusion process, Eq. (172), is in thermal equilibrium.

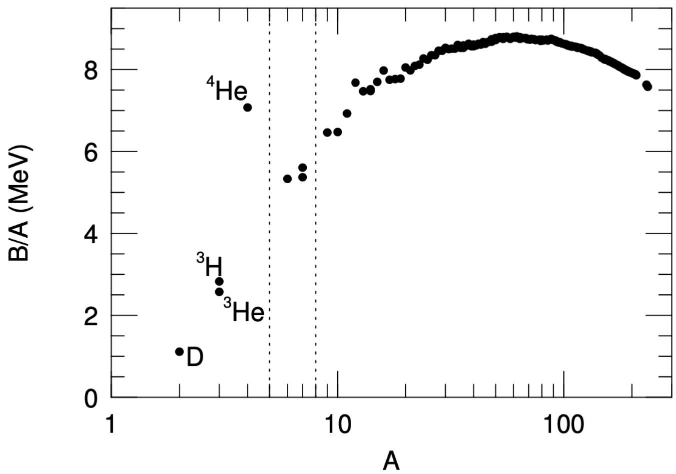
Based on this, we can compute the ratio n_D/n_p, by moving the factor n_n to the right-hand side, and using $n_n \simeq (1/6)\eta_b n_\gamma$ (valid up to decay and fusion consumption effects) and $n_\gamma = 2\zeta(3)T^3/\pi^2$ (see Eq. (135) with $g_\gamma = 2$),

$$\frac{n_D}{n_p} \simeq \eta_b n_\gamma \left(\frac{\pi}{m_p T} \right)^{3/2} e^{\Delta/T} = \eta_b \frac{2\zeta(3)}{\sqrt{\pi}} \left(\frac{T}{m_p} \right)^{3/2} e^{\Delta/T} \ . \tag{178}$$

A plot of this ratio as a function of photon temperature is shown below. The curve is valid (actually quite accurate) only up to a point before the ratio reaches 1/4 (the maximum). Physically, the processes that eventually stop the deuteron population from further growing include the formation of heavier elements (in particular Helium) and the decay of neutron. Both effects contribute extra suppression factors to the right-hand side of the above equation.



As a further step, let's open our mind to the presence of more than one nucleus. The plot below shows the binding energy per nucleon for various nuclei. Clearly, among the light elements, helium 4 is special for its relatively highest binding energy per nucleon. It is energetically favored for He-4 to be formed compared to other nuclei such as deuteron (D), tritium (H-3), He-3, etc.



In order to get to ^{4}He, we need a series of processes that pass through lighter nuclei.

$$
\begin{aligned}
D + p &\to {}^3\text{He} + \gamma \\
D + n &\to {}^3\text{H} + \gamma \\
D + D &\to {}^4\text{He} + \gamma \\
{}^3\text{He} + n &\to {}^4\text{He} + \gamma \\
{}^3\text{H} + p &\to {}^4\text{He} + \gamma \\
&\cdots
\end{aligned}
\tag{179}
$$

Therefore, as the population of $D,^3$H$,^3$He are built up to some extent, they quickly go on to form ^{4}He. This explain why in the following picture their populations first grow but then drops. They sacrifices themselves to form ^{4}He which is the most abundant nuclei after hydrogen in the universe today.

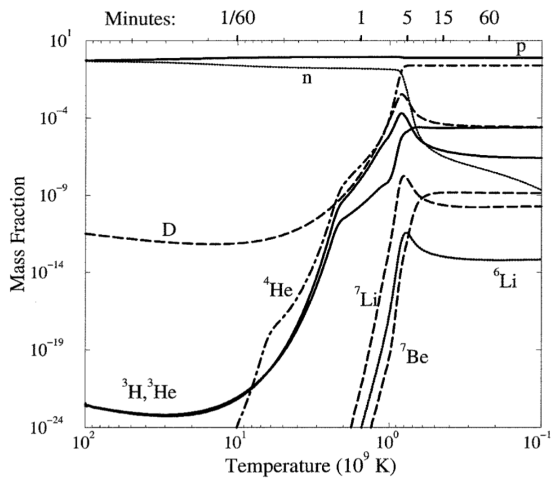
The fusion process can continue and go beyond ^{4}He, allowing heavier nuclei to be formed, such as lithium.

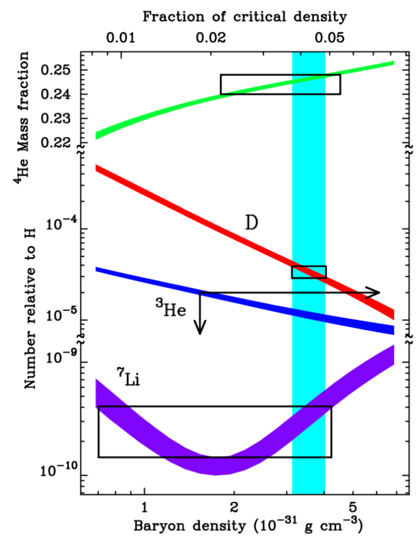
To calculate the formation rate of all the light elements, we can apply the Saha equation to nuclei with Z protons and $A - Z$ neutrons, if the fusion process is in equilibrium. After the fusion drops out of equilibrium (correspondingly $\Gamma < H$), the number densities of the elements stop evolving but simply dilutes with the expansion of the universe.

To actually compute the BBN processes accurately, we need to write down a number of Saha equations coupled together, one for each element. One usually needs a code to solve them. We do not resort to details here, but simply present the plot below (left panel) which is the result of a numerical calculation.

From the left plot, we also notice that the neutron number density does not decay exponentially, because they can be recreated during the fusion processes of heavy elements, such as $^3\text{H} + D \to {}^4\text{He} + n$. It will eventually drop exponentially as e^{-t/τ_n} after BBN is completed.

The resulting abundance of Helium-4 (and that of other light elements) is measurable and is closely related to cosmological parameters. See the right panel. Mention ^{7}Li anomaly.





BBN ends at temperature around $1 - 10\,\text{keV}$ scale, where the age of the universe was about $10^4 - 10^6$ sec (within a day). Around this time, all free neutrinos have been exhausted. Interactions among the remaining nuclei finally stops because they are all positively charged under electromagnetism and repulsive to each other.

17 Recombination

Let us now dial the clock to a time well after BBN, where the age of the universe is about 10^{13} sec $\sim$ 300,000 years ($T_\gamma \sim$ a few $\times\, 0.1\,\text{eV}$). Both electrons and protons (as well as other nuclei) are non-relativistic by then. This is the time when neutral atoms are about the form – recombination. This is also the epoch when the cosmic microwave background (CMB) is formed.

Clearly the most important atom is the hydrogen atom. Before this time, the universe is too hot with too energetic photons that can easily break atoms apart. Recall the binding energy of ground state hydrogen is $BE = 13.6\,\text{eV}$. (Here is a subtlety: the helium atom has a higher binding energy, thus they actually form earlier than the hydrogen. We do not worry about them because the protons are much more abundant than the helium nuclei in the universe.)

The key process for recombination is

$$e^- + p^+ \to H + \gamma \ . \tag{180}$$

Electromagnetic interaction is very fast and this process remains in thermal equilibrium until most of electrons and protons in the universe end up in atoms. We can write down a similar Saha equation as Eq. (176),

$$\frac{n_H}{n_e^2} \simeq \frac{n_H}{n_e n_p} = \frac{g_H}{g_p g_e}\left(\frac{m_H}{m_e m_p}\right)^{3/2}\left(\frac{T}{2\pi}\right)^{-3/2} e^{BE/T} \simeq \left(\frac{2\pi}{m_e T}\right)^{3/2} e^{BE/T} \ . \tag{181}$$

where we have used the approximation $m_p \simeq m_H$. There are two types of hydrogen atom, with total spin 1 and 0. In together, they contribute to $g_H = 4$. Electron and proton are spin 1/2 fermions, thus $g_e = g_p = 2$.

Eq. (181) is a very useful equation because the right-hand side is only a function of T. The left-hand side involves two number densities. n_H is the number density of hydrogen atom that has been formed, which is equal to the number density of electrons bounded inside atoms. n_e is the number density of free electrons that are not bounded in atoms. Clearly, the sum $n_H + n_e$ is equal to the total number density of electrons in the universe. Let us forget about the existence of helium or other heavy atoms as an approximation (they are less abundant anyway). Assuming the universe is neutral, $n_H + n_e$ is also equal to the total number density of protons, or baryons approximately. This leads to a constraint

$$n_e + n_H = \eta_b n_\gamma \ , \tag{182}$$

where $\eta_b = 6 \times 10^{-10}$ is the cosmic baryon asymmetry introduced in Eq. (164).

Let's define a useful squantity called the **ionization fraction**,

$$X_e = \frac{n_e}{n_e + n_H} \ . \tag{183}$$

Clearly, before recombination $X_e = 1$ because $n_H = 0$. After the completion of recombination, $X_e \simeq 0$. A bit algebra shows

$$\frac{1 - X_e}{X_e^2} = \frac{n_H}{n_e^2}(n_e + n_H) \simeq \frac{n_H}{n_e^2}\eta_b n_\gamma \ . \tag{184}$$

Together with Eq. (181), we can derive

$$\frac{1 - X_e}{X_e^2} = \eta_b \frac{2\zeta(3)}{\pi^2}\left(\frac{2\pi T}{m_e}\right)^{3/2} e^{BE/T} \ . \tag{185}$$

This equation satisfies the correct behavior mentioned above. At early time (large T), we have $X_e \simeq 1$ which makes left-hand side very small. Correspondingly, the right-hand side of the equation is highly suppressed by the small parameter η_b. At late time (small T), the right-hand side is very large due to the exponential factor. This implies $X_e \simeq 0$. With Eq. (185), we can derive solve for X_e at any time for any temperature T.

We can define a special time when $X_e = 0.1$, i,.e., 90% of the electrons have formed neutral hydrogen. This correspond to a temperature which can be found by numerically solving Eq. (185),

$$T_{\rm rec} = 0.296\,{\rm eV} = 3435\,{\rm K} \ . \tag{186}$$

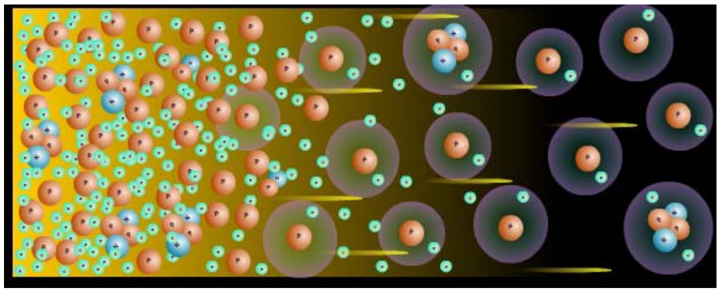
This is defined as the recombination temperature.

Because the photon temperature simply scales as $1/a$ between recombination and today, and today's CMB temperature is $T_0 = 2.7\,{\rm K}$, we can derive the scale factor at time of recombination is $a_{\rm rec} = T_0/T_{\rm rec} = 7.86 \times 10^{-4}$. The corresponding redshift factor is $z_{\rm rec} = (1/a_{\rm rec}) - 1 \simeq 1272 \sim 1300$. The number 1300 is commonly quoted in textbooks and literatures. We can compute the corresponding age of the universe at time of recombination using Eq. (123),

$$t_{\rm rec} = \frac{1}{H_0}\int_0^{a_{\rm rec}} \frac{da'}{\sqrt{\Omega_\Lambda^0 a'^2 + \frac{\Omega_m^0}{a'} + \frac{\Omega_\gamma}{a'^2}(1+r)}} = \frac{2.1 \times 10^{-5}}{H_0} = 9 \times 10^{12}\,{\rm sec} \simeq 3 \times 10^5\,{\rm years} \ . \tag{187}$$

where we used $H_0 = 70\,{\rm km/s/Mpc} = 1/(4.4 \times 10^{17}\,{\rm sec})$. The factor $r = \rho_\nu/\rho_\gamma = (7/8) * 3 * (4/11)^{4/3} = 0.68$ takes into account of neutrino contribution to radiation energy density. (Numerically, it is actually not necessary to be so careful because the universe has already entered the matter-dominated era. Note: $z_{\rm eq} \simeq 3400 > z_{\rm rec}$.)

To derive the above recombination temperature and time, we focused on the population of free electrons or that of the hydrogen in the universe. Clearly, when the quantity X_e becomes sufficiently small, the universe becomes mainly comprised of neutral atoms instead of ions (free electrons and protons, i.e., charged particles), and become transparent to photons. The latter could then free stream across the universe and reach our telescope and be observed as the CMB. See cartoon picture below.



Now we put our focus on the photons and derive the time when they have the last scattering with the ions. Afterwards, they (or the majority of them) are free from further interactions. The time of last scattering is defined as $\Gamma = n_e \sigma v = H$, where σ is the electro-magnetic interaction between photon and the background electron. Here we simply assume the low-energy photons do not interact with hydrogen atom at all. For non-relativistic electron, this is well described by the Thomson scattering and the cross section is

$$\sigma = \frac{8\pi\alpha^2}{3m_e^2} \ , \tag{188}$$

where $\alpha = 1/137$ is the fine-structure constant.

The Hubble parameter around the time of recombination can be calculated using the Friedmann equation

$$H(T) = H_0\sqrt{\Omega_\Lambda^0 + \frac{\Omega_m^0}{a'^3} + \frac{\Omega_r}{a'^4}(1+r)} = H_0\sqrt{\Omega_\Lambda^0 + \Omega_m^0\left(\frac{T}{T_0}\right)^3 + \Omega_\gamma\left(\frac{T}{T_0}\right)^4(1+r)} \ . \tag{189}$$

The value of H_0 is 1.5×10^{-42} in unit of GeV. The photon temperature today is $T_0 = 2.7\,\mathrm{K}$.

The electron number density can be calculated using $n_e(T) = \eta_b n_\gamma(T) X_e(T)$, where $X_e(T)$ can be solved using Eq. (185). With the above temperature dependent functions, we can write down the decoupling condition equation $n_e\sigma v = H$ as

$$\eta_b\left(\frac{2\zeta(3)}{\pi^2}T^3\right)X_e(T)\left(\frac{8\pi\alpha^2}{3m_e^2}\right) = H(T) \ . \tag{190}$$

The resulting solution for temperature (last scattering) is found to be

$$T_{\rm ls} = 0.265\,\mathrm{eV} = 3075\,\mathrm{K} \ . \tag{191}$$

This temperature corresponds to a scale factor $a_{\rm ls} = T_0/T_{\rm ls} = 8.8\times10^{-4}$. The corresponding redshift factor is $z_{\rm ls} = (1/a_{\rm ls}) - 1 \simeq 1138 \sim 1100$. The corresponding age of the universe is

$$t_{\rm ls} = \frac{1}{H_0}\int_0^{a_{\rm ls}} \frac{da'}{\sqrt{\Omega_\Lambda^0 a'^2 + \frac{\Omega_m^0}{a'} + \frac{\Omega_\gamma}{a'^2}(1+r)}} = \frac{2.34\times10^{-5}}{H_0} = 1.1\times10^{13}\,\mathrm{sec} = 3.5\times10^5\,\mathrm{years} \ . \tag{192}$$

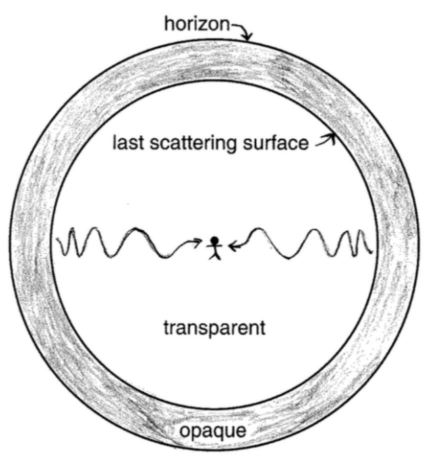
Here is the final comment of this section. Because the universe is homogeneous, we can imagine that the whole universe suddenly becomes transparent to photons at time $t_{\rm ls}$. As a result, photons become free particles and continue traveling with their original momentum set by the thermal distribution. The releasing of background photons happens everywhere at the same time.

It is important to appreciate that today we are only observing the CMB at a particular point in the universe. We cannot make observations anywhere else. This means we are only seeing part of the CMB photons – whose originating from the surface of a sphere with a radius $r_{\rm ls}$. We are at the center of the sphere. This surface is called the **last scattering surface**. Photons originating from any point inside the last scattering surface have already passed our position and cannot be observed. Photons originating from points outside the last scattering surface cannot be released because the universe was opaque at early times.

The radius $r_{\rm ls}$ today can be calculated using the familiar formula

$$r_{\rm ls} = \int_{t_{\rm ls}}^{t_0} \frac{dt}{a(t)} = \frac{1}{H_0} \int_{a_{\rm ls}}^{1} \frac{da}{\sqrt{\Omega_\Lambda^0 a^4 + \Omega_m^0 a + \Omega_\gamma(1+r)}} = \frac{3.12}{H_0} \ . \tag{193}$$

Recall the particle horizon of the universe today is about $3.2/H_0$ (see Eq. (129)), which corresponds to the same integral but with the lower limit of t integral equal to 0. The last scattering surface is only slightly smaller than the particle horizon today. Below is a good picture of it.



18 Dark Matter

In earlier discussion, we have mentioned that dark matter plays the key role of generating inhomogeneities in our universe. In this section, we first take a closer look at how this works. Afterwards, we will briefly discuss several other evidence for the existence of dark matter.

Jeans instability. Because we do not know wether dark matter is made of elementary particles or not, we will move back to the fluid language. In general, there are three quantities used to describe the evolution of an inhomogeneous fluid:

$$\rho(\vec{r}, t), \quad p(\vec{r}, t), \quad \vec{v}(\vec{r}, t) \ . \tag{194}$$

In addition to energy density and pressure, we also need the velocity of the fluid $\vec{v}$. In the homogeneous limit, $\rho(\vec{r},t) \to \rho(t)$, $p(\vec{r},t) \to p(t)$, and $\vec{v}(\vec{r},t) \to 0$.

There are two fundamental equations for fluid dynamics. For a free fluid with Minkowski spacetime,

- Continuity equation:
$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j} = 0 \ , \tag{195}$$
where $\vec{j} = \rho \vec{v}$.

- Euler equation:
$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla})\vec{v} = -\frac{\vec{\nabla} p}{\rho} \ . \tag{196}$$
In this equation, the gradient of pressure plays the role as a classic "force" that leads to acceleration.

Next, we apply these equations to the matter fluid in nearly homogeneous the universe, by writing

$$\begin{aligned} \rho(\vec{r},t) &= \bar{\rho}_0(t) + \delta\rho(\vec{r},t) \ , \\ p(\vec{r},t) &= \bar{p}_0(t) + \delta p(\vec{r},t) \ , \\ \vec{v}(\vec{r},t) &= \vec{0} + \delta\vec{v}(\vec{r},t) \ , \end{aligned} \tag{197}$$

where the first term in each expansion describes a perfectly homogeneous universe (without $\vec{r}$ dependence), and the second term is the perturbation up to linear order.

Plugging the above expansions into the continuous and Euler equations, and keep the terms linear in the small perturbative quantities, we get

$$\begin{aligned} \frac{\partial \delta\rho}{\partial t} + \bar{\rho}\vec{\nabla} \cdot \delta\vec{v} = 0 \ , \\ \frac{\partial \delta\vec{v}}{\partial t} = -\frac{\vec{\nabla}\delta p}{\bar{\rho}} \ . \end{aligned} \tag{198}$$

These two equations can be combined into a second-order differential equation for ρ_1,

$$\frac{\partial^2 \delta\rho}{\partial t^2} - \nabla^2 \delta p = \left(\frac{\partial^2}{\partial t^2} - c_s^2 \nabla^2\right) \delta\rho = 0 \ , \tag{199}$$

where in the second step we introduced the speed of sound $c_s^2 \equiv \delta p/\delta\rho$. The above $\delta\rho_1$ equation features plane-wave solutions of the form

$$\delta\rho = A \sin(\vec{k} \cdot \vec{r} - \omega t + \delta_1), \quad \omega = c_s k \ , \tag{200}$$

where $k = |\vec{k}|$.

As the next step, we must include the gravitational effect between the density perturbation ρ_1 and the background. For non-relativistic matter dominated universe, it is easy to add this effect by introducing a gravitational force term to the right-hand side of the Euler equation. Eq. (196) now becomes

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla})\vec{v} = -\frac{\vec{\nabla} p}{\rho} - \vec{\nabla}\Phi \ , \tag{201}$$

where Φ is the gravitational sourced by ρ and satisfies the Poisson equation

$$\nabla^2\Phi = 4\pi G_N \rho \ . \tag{202}$$

Here we assume the universe has evolved to the stage of matter domination, thus ρ is given by the energy density of matter. If we repeat the perturbation approach by plugging Eq. (197) into the continuous equation (194) and the new Euler equation (201), we will get

$$\begin{aligned} \frac{\partial \delta\rho}{\partial t} + \bar{\rho}\vec{\nabla}\cdot\delta\vec{v} = 0 \ , \\ \frac{\partial \delta\vec{v}}{\partial t} = -\frac{\vec{\nabla}\delta p}{\bar{\rho}} - \vec{\nabla}\delta\Phi \ , \end{aligned} \tag{203}$$

where $\delta\Phi$ is the leading perturbation in the gravitational potential, and satisfies $\nabla^2\delta\Phi = 4\pi G_N \delta\rho$. The resulting second order differential equation for $\delta\rho$ now becomes

$$\frac{\partial^2\delta\rho}{\partial t^2} - \left(c_s^2\nabla^2 + 4\pi G_N\bar{\rho}\right)\delta\rho = 0 \ . \tag{204}$$

Plugging in a plane wave solution, we find a new dispersion relation

$$\omega = \pm\sqrt{c_s^2 k^2 - 4\pi G_N \bar{\rho}} \ . \tag{205}$$

The important physics point here is that for sufficiently small k (large length scales), ω will become imaginary. The corresponding phase factor $e^{-i\omega t}$ will exponentially grow with time. As a result, the perturbation ρ_1 grows exponentially with time, leading to more and more inhomogeneous universe. It serves as the seed for forming the large scale structure of the universe.

The Jeans length for structure is defined as

$$\lambda_J = \frac{2\pi}{k_J} = c_s\sqrt{\frac{\pi}{G_N\bar{\rho}}} \ . \tag{206}$$

We have not discussed much about the sound speed c_s yet. For non-relativistic particles with a temperature, it is equal to $\sqrt{5T/(3m)} \ll 1$. This is the place where dark matter and atomic matter make a difference. Because dark matter does not couples to the photons, the sound speed of dark matter is much lower than that of atoms. According to Eq. (205), it is thus much easier for dark matter to form inhomogeneous structures at small distances.

The Jeans swindle.
There is actually an inconsistency in the above derivations. At zeroth order, Eq. (201) can written as

$$\vec{\nabla}\bar{\Phi} = 0 \ , \tag{207}$$

where we neglect all perturbations, and set $\vec{v} \to 0$ and used $\bar{p}$ is homogeneous thus $\vec{\nabla}\bar{p} = 0$. This then implies

$$\nabla^2\bar{\Phi} = 0 \ , \tag{208}$$

which is clearly in contradiction with the Poisson equation (202), which states $\nabla^2\bar{\Phi} = 4\pi G\bar{\rho}$. The fact that we reached the above argument for inhomogeneity growth by neglecting this inconsistency is sometimes referred to the *Jeans swindle*. This problem will be solved when we put the fluid in an expanding universe (we have not done so yet). Leave as assignment.

To explore further on structure formation and the corresponding observables, see David Tong's lecture note: `http://www.damtp.cam.ac.uk/user/tong/cosmo/three.pdf`.

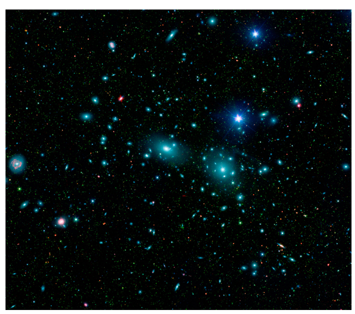
***** If time permits, add expansion rate, solve Jean's swindle, derive evolution of perturbations and matter power spectrum – see handwritten note.**

Other important evidence for dark matter

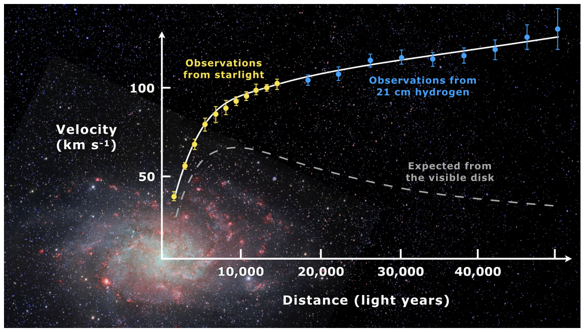
Neptune. (not dark matter) Historically, planet Neptune was inferred to exist because of the unexpected changes in the orbit of Uranus, as observed by astronomers. The discovery of dark matter is also based on experimental evidence and follows a similar way of thinking.

As we did in this course, we always assume the theory of gravity (Einstein) is correct. Alternative proposals to dark matter include modified gravity (MOND, stands for MOdified Newtonian Dynamics). However, no single such theory is known to exist that explains the experimental evidence for dark matter on all the length scales and remains consist with other cosmological data. General relativity has passed many tests and serves as an important pillar for modern physics.

Coma Cluster. The Coma Cluster is a large cluster of galaxies that contains over 1000 identified galaxies. Back in 1933, Fritz Zwicky, a professor at Caltech, showed that the galaxies of the Coma Cluster were moving too fast for the cluster to be bound together by the visible matter of its galaxies. Though the idea of dark matter would not be accepted for another fifty years, Zwicky wrote that the galaxies must be held together by some *dunkle Materie.*



Galactic Rotational Curves. In the late 1960s and early 1970s, Vera Rubin, an astronomer at the Carnegie Institution of Washington, worked with a new sensitive spectrograph and measured the velocity curve of edge-on spiral galaxies (e.g., the Andromeda galaxy) to a greater degree of accuracy than had ever before been achieved. Discovery was announced in 1975 that that most stars in spiral galaxies orbit at roughly the same speed. This implies that galaxy masses grow approximately linearly with radius well beyond the location of most of the stars (the galactic bulge). See figure right below (rotation curve of spiral galaxy Messier 33).



If we only take into account of the gravitational force due to the visible matter (stars and gas) which are located near the center of the galaxy, at large enough distance, we would derive (Newton's law works effectively well here)

$$\frac{GM_{\text{visible}}}{r^2} = \frac{v(r)^2}{r} \ , \tag{209}$$

where M_{visible} is constant, and thus

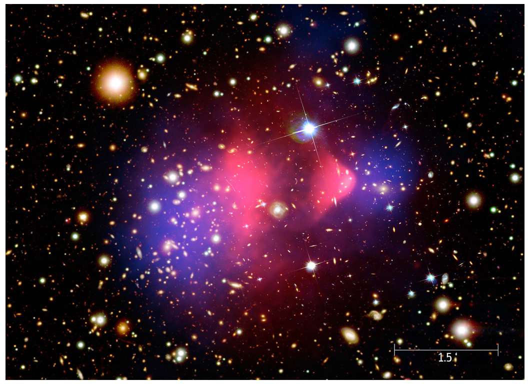
$$v(r) \sim \sqrt{1/r} \ . \tag{210}$$

which is in contradiction with observation. The gravitating mass has to be larger than M_{visible} and grows with r.

Gravitational Lensing. Galaxy clusters are the largest gravitationally bound structures in the Universe which can deflect light-rays traveling near them. As seen from Earth, this effect can distort the image of background source object as multiple images, arcs, and rings (cluster strong lensing). More generally, the effect causes small, but statistically coherent, distortions (cluster weak lensing). See figure right below, which is an image taken by the Hubble Space Telescope.

Bullet Cluster. An important application of the gravitational lensing effect is in the observation of the bullet cluster event.

In the figure right below (galaxy cluster 1E 0657-56, also known as the bullet cluster), the image in background showing the visible spectrum of light taken by the Magellan temescopes and Hubble Space Telescope. The blue overlay represents the mass distribution of the clusters detived from gravitational lensing effects. The pink overlay shows the x-ray emission (recorded by the Chandra Telescope) of the colliding clusters.



The clear separation between normal atomic matter and dark matter has not been seen before and gives the strongest evidence yet that most of the matter in the Universe is dark.

So what is dark matter's identity?

This is a million dollar question and we do not know the answer yet. There are many dark matter candidates ranging from massive black holes to elementary particles. If any of them makes up the dark matter in the universe ($\sim 25\%$ of total energy budget), we know it has a population in the universe, depending on its mass. An important physics question is then can we find this population of dark matter in our laboratories, without relying on the gravitational effects mentioned above. If yes, this would be a great independent test. This often requires dark matter to have some interaction with known matter beyond gravitational forces.

As an example, the particle physics experimental group at Carleton is a member of the DEAP experiment at SNOLab. The experiment is made of liquid argon and targets on an important class of dark matter candidates (WIMP) which has a mass in the range of 1-1000 GeV and participates in the weak interaction. Theoretically, it is very appealing for WIMP to be dark matter.

There are a number of other theoretically appealing dark matter candidates that require different search methods.

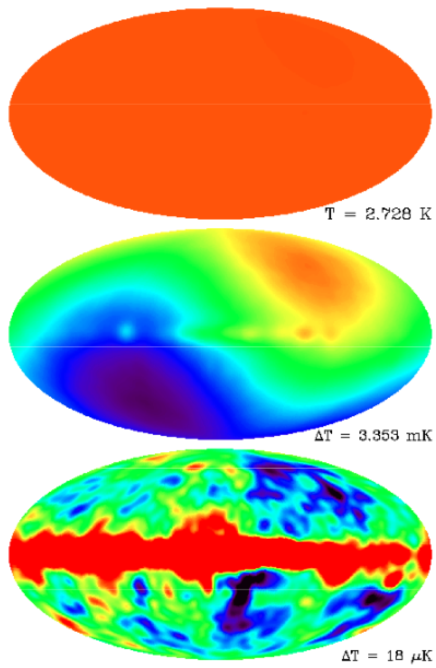
So far, none of the dark matter candidates has been found in the laboratory yet.

Identifying the nature of dark matter has served as one of the most exciting and ongoing research directions in the field of astroparticle physics for the past decade (and will continue to do so for the near future, or until we find the answer :).

19 CMB anisotropy

With the above discussion of inhomogeneous universe and structure formation, it is not hard to imagine that the mapping of inhomogeneity on the last scatter surface will manifest as anisotropy.

The CMB photon temperature can be measured as the following. Assuming it is a blackbody radiation, the spectral radiance (total photon energies passing your telescope per unit area per unit time per unit solid angle) is $B(\nu, T) = (2h\nu^3/c^2)/(e^{h\nu/(kT)} - 1)$. Thus, if we measure the value of B at certain frequency, we can infer the corresponding photon temperature. If we measure the values of B at various frequencies, we can verify that the Planck's law of ν distribution. For the temperature measurement to be done, we do not have to look around the whole sky. It is sufficient to take data from a small patch of the sky. The smallest size of the patch depends on the angular resolution of your telescope. As a result, we can make many measurement of the CMB temperature that correspond to various patches viewed by us at various angular directions. Using spherical coordinate (we sit at the origin), these angular directions are labelled by different values of θ and ϕ. As a result, the measured T is a function of θ, ϕ.



Because the universe is approximately isotropic (and homogeneous), the θ, ϕ dependence in T should be very mild. This is indeed observed by the first measurement of CMB done in the history. Averaging over the whole sky, it was found today $T_0 = 2.7\,\mathrm{K}$.

The precision of experiment is related to how well you can tell two slightly different temperatures (ΔT) apart. With higher and higher precision, it has been observed (famously first by the COBE experiment) that CMB photon temperature is not strictly isotropic from all directions. First, by reaching the precision $\Delta T \sim$ mK (milli-Kelvin), the temperature distribution exhibits a dipole. This can be explained because we are not exactly free-falling observers, but instead moving along with the galaxy cluster that we belong to at the velocity $v = 369 \pm 3\,\mathrm{km/s}$. This velocity is much smaller than c but large enough to induce a true Doppler effect, with shifted frequency as large as $\Delta\nu \simeq (v/c)\nu \sim 10^{-3}\nu$. For given measured B, "wrong" value of ν will lead to wrong inferred temperature. Roughly, $\Delta T/T \sim \Delta\nu/\nu$. For $T = T_0 = 2.7\,\mathrm{K}$, this leads to $\Delta T \sim$ mK, as is measured. We can remove this dipole feature in the $T(\theta, \phi)$ distribution by changing to the free-fall frame.

The state-of-art CMB experiments have reached very high precision in the temperature measurement. They find that the CMB temperature distribution exhibits stochastic fluctuation at $\sim 10\mu$K level, corresponding to $\Delta T/T \sim 10^{-5}$.

Let's be more quantitative on what the fluctuation means. Define the deviation of temperature from the average value T_0 as $\delta T \equiv T(\theta, \phi) - T_0$. Clearly, δT is also a direction $\hat{n} = (\theta, \phi)$ dependent quantity. If we simply average δT in all directions, we will get zero.

Let's expand δT in terms of spherical harmonics

$$\delta T(\hat{n}) = T_0 \sum_{\ell,m} a_{\ell m} Y_{lm}(\hat{n}) \ . \tag{211}$$

Using the normalization of spherical harmonics, we get

$$a_{\ell m} T_0 = \int d\hat{n} \delta T(\hat{n}) Y^*_{lm}(\hat{n}) \ . \tag{212}$$

The average over all directions gives

$$\begin{aligned}
\langle \delta T \rangle &= \frac{1}{4\pi} \int d\hat{n} T_0 \sum_{\ell,m} a_{\ell m} Y_{lm}(\hat{n}) \\
&= \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta T_0 \sum_{\ell,m} a_{\ell m} Y_{lm}(\theta,\phi) \\
&= \frac{1}{4\pi} T_0 a_{00} \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta Y_{00}(\theta,\phi) \\
&= \frac{1}{\sqrt{4\pi}} T_0 a_{00} \ .
\end{aligned} \tag{213}$$

This is zero by definition and implies $a_{00} = 0$.

We can get nonzero averages by considering two-point correlation functions, defined as

$$\begin{aligned}
\Big\langle \delta T(\hat{n}_1) \delta T(\hat{n}_2) \Big\rangle_{\hat{n}_1 \cdot \hat{n}_2 = \cos\Theta} &\equiv \frac{\int d\hat{n}_1 \int d\hat{n}_2 \delta T(\hat{n}_1) \delta T(\hat{n}_2) \times \delta(\hat{n}_1 \cdot \hat{n}_2 - \cos\Theta)}{\int d\hat{n}_1 \int d\hat{n}_2 \delta(\hat{n}_1 \cdot \hat{n}_2 - \cos\Theta)} \\
&= \frac{1}{8\pi^2} \int d\hat{n}_1 \int d\hat{n}_2 \delta T(\hat{n}_1) \delta T(\hat{n}_2) \times \delta(\hat{n}_1 \cdot \hat{n}_2 - \cos\Theta) \ ,
\end{aligned} \tag{214}$$

where the last factor in the integrand is a Dirac-δ function. We hold the relative angle between $\hat{n}$ and $\hat{n}'$ fixed and average over $\hat{n}$ and $\hat{n}'$. Experimentally, the averaging procedure is first take a small patch p_1 from the sky, find all the other patches p_{1j} at directions with relative angle θ with respect to p_1, and find each $\delta T(p_1)\delta T(p_{1j})$. And then we move on to other patches $p_2, p_3, \cdots$ throughout the sky. Finally we find the average value of all the products in the set $\{\delta T(p_i)\delta T(p_{ij})\}$.

Clearly, the correlation function defined Eq. (214) is a function of $\cos\Theta$, we can carry out a partial wave expansion,

$$\Big\langle \delta T(\hat{n}_1) \delta T(\hat{n}_2) \Big\rangle_{\hat{n}_1 \cdot \hat{n}_2 = \cos\Theta} = \frac{1}{4\pi} \sum_\ell (2\ell+1) C_\ell P_\ell(\cos\Theta) \ , \tag{215}$$

where $P_\ell(x)$ is the Legendre polynomial. Using the normalization property of P_ℓ,

$$\int_{-1}^{1} dx P_\ell(x) P_{\ell'}(x) = \frac{2}{2\ell+1} \delta_{\ell\ell'} \ , \tag{216}$$

we can write C_ℓ as

$$C_\ell = 2\pi \int_{-1}^{1} d\cos\theta \Big\langle \delta T(\hat{n}_1) \delta T(\hat{n}_2) \Big\rangle_{\hat{n} \cdot \hat{n}' = \cos\Theta} P_\ell(\cos\Theta) \ . \tag{217}$$

Plugging Eq, (214) into the integrand on the right-hand side, and annihilating the $\int d\cos\theta$ integral with the δ-function, we get

$$C_\ell = 2\pi \frac{1}{8\pi^2} \int d\hat{n}_1 \int d\hat{n}_2 \delta T(\hat{n}_1) \delta T(\hat{n}_2) P_\ell(\hat{n}_1 \cdot \hat{n}_2) \ . \tag{218}$$

Next, we apply the identity

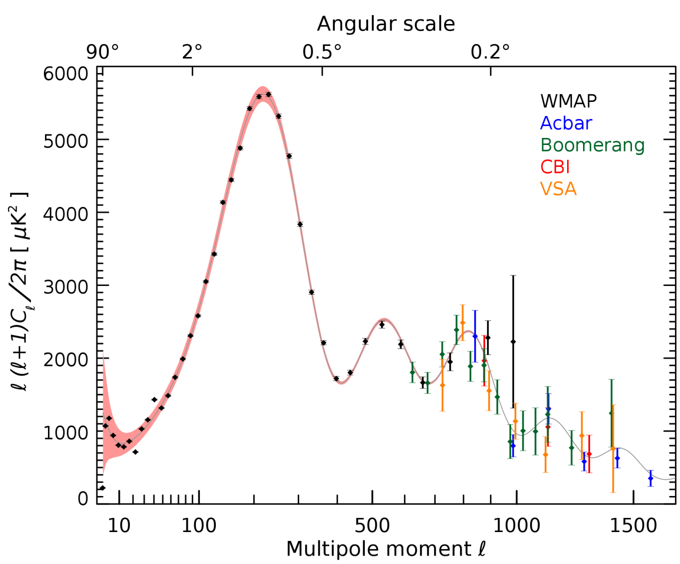
$$P_\ell(\hat{n}_1 \cdot \hat{n}_2) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} Y^*_{\ell m}(\hat{n}_1) Y_{\ell m}(\hat{n}_2) \ . \tag{219}$$

This leads to

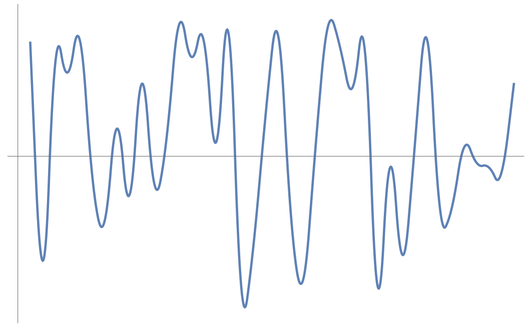
$$\begin{aligned} C_\ell &= \sum_{m=-\ell}^{\ell} \frac{8\pi^2}{2\ell+1} \frac{1}{8\pi^2} \int d\hat{n}_1 \delta T(\hat{n}_1) Y^*_{\ell m}(\hat{n}_1) \int d\hat{n}_2 \delta T(\hat{n}_2) Y_{\ell m}(\hat{n}_2) \\ &= \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 T_0^2 \ . \end{aligned} \tag{220}$$

In the second step, we used Eq. (212).

The experimentally measured value of C_ℓ is given in the figure below. Clearly, there are many peaks at various angles. They are called acoustic peaks. The peak implies strong angular correlation between two directions separated by certain angles.



The CMB anisotropy originates from the spatial inhomogeneity (3D) of the stuff distributed in the universe. The last scattering of photons occur at $a_{\rm ls} = 8.8 \times 10^{-4}$. Recall that matter-radiation equality occurs at $a_{\rm eq} = 1.7 \times 10^{-4} < a_{\rm ls}$. This comparison implies that the universe has already enter the matter-dominated era when CMB forms. As a result, matters (atoms and dark matter) leave important imprints on the CMB spectrum. It is beyond the scope of this course to systematically work out the cosmic perturbation theory for inhomogeneities and anisotropies. Instead, we just give a taste of the relevant physics by presenting an intuitive explanation of the first (and largest) peak in the CMB TT-spectrum as shown above.



The key starting point is that our universe is not strictly homogeneous. In early universe, the total energy density has small primordial fluctuations at the level of $\pm 10^{-5}$ around the central value. A 1D random fluctuation is depicted in the picture above. These small fluctuations are the seeds for forming structures (clusters, galaxies, etc) in the later stage of the universe. Around the time of CMB formation, matter already has dominated the energy density of the universe. Among the matter species, about 80% is dark matter and only 20% is made of baryons (and electrons). In the more dense regions (corresponding to peaks in the above random distribution), dark matter want to concentrate even more resulting in magnitude of fluctuation much larger than 10^{-5}. This is due to the Jeans instability to be discussed in the next section. In contrast, baryons are still ionized before recombination and are strongly coupled to photons. As a result, they feel stronger and stronger radiation pressure when falling into the over-dense regions of the universe. Eventually, a sound wave is built up propagating in the radial directions away from the center of over-dense region. (This is called **baryon acoustic oscillation**.) Sound waves are waves packet containing over density of baryons. They travel for a finite distance until the last scattering of photons occur where they run out of the fuel of pressure. This leads to a characteristic length scale, which on the last scattering surface corresponds to rings (see picture below for illustration), and a characteristic angular separation. This length scale l today can be estimated as

$$l = a(t_0) \int_0^{t_{\rm ls}} \frac{c_s dt}{a} = c_s \int_0^{a_{\rm ls}} \frac{da}{Ha^2} = \frac{c_s}{H_0} \int_0^{a_{\rm ls}} \frac{da}{\sqrt{\Omega_\Lambda a^4 + \Omega_m a + \Omega_\gamma (1+r)}} = \frac{0.069}{H_0} \frac{c}{\sqrt{3}} \ . \tag{221}$$

The lower limit of the time or a integral is not important. In the last step, we estimate the sound speed with $c_s \approx c/\sqrt{3}$.

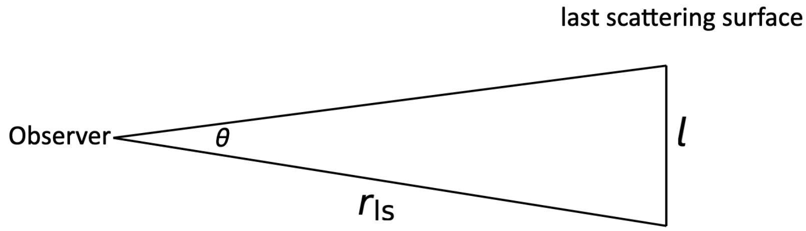
On the other hand, this characteristic length scale is observed by us today, at a distance of $r_{\rm ls}$ as calculated in Eq. (193). As a result, the angular spanned by the length scale L in view of the observer today is (see the picture below)

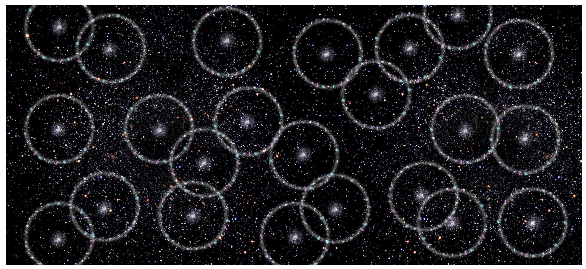
$$\theta = \frac{l}{r_{\rm ls}} \simeq 0.02 \simeq 1^\circ \ . \tag{222}$$

Here we have assumed that the universe is flat otherwise the estimate of θ will depend on the curvature of the universe. This angle corresponds to a multipole moment given by

$$\ell \sim \frac{\pi}{\theta} \sim 200 \ . \tag{223}$$

This estimate gives roughly the correct position of the first acoustic peak in the CMB anisotropy spectrum.





20 Inflation

In this section, let's raise several philosophical questions on the overall consistency between the universe we observe and the standard big-bang theory, based on some rather simple consideration. As we will see, a profound idea can follow in this approach.

The horizon problem.
Back in Section 12, we have calculated the particle horizon of the universe today is equal to $d_H(t_0) = 3.19/H_0$. We have also calculated the time dependence of $d_H(t)$ which grows monotonically with time. Recall that particle horizon is defined as the distance light can travel since the very beginning of time, thus a region in the universe with radius $d_H(t)$ defines the largest causally connected patch at given time t. The fact $d_H(t_0)$ is much larger than $d_H(t)$ in the far past implies that our presently causally connected universe is made of many small patches of the past. At time t, each patch was causally connected but no causal contact had been established among any two patches yet. In principle, each patch could have been quite different from another (in temperature, energy densities, etc). This naturally leads to question: why is our universe today so homogeneous?

Let's further sharpen this question by making an apple-to-apple comparison. At the time of last scattering when CMB was born, the particle horizon of the universe was

$$
\begin{aligned}
d_H(t_{\rm ls}) &= a(t_{\rm ls})\int_0^{t_{\rm ls}} \frac{dt}{a(t)} = a(t_{\rm ls})\int_0^{a_{\rm ls}} \frac{da}{a(t)^2 H(t)} = \frac{a(t_{\rm ls})}{H_0}\int_0^{a_{\rm ls}} \frac{da}{a^2\sqrt{\Omega_\Lambda + \Omega_m a^{-3} + \Omega_\gamma(1+r)a^{-4}}} \\
&= \frac{a(t_{\rm ls})}{H_0}\int_0^{a_{\rm ls}} \frac{da}{\sqrt{\Omega_\Lambda a^4 + \Omega_m a + \Omega_\gamma(1+r)}} = \frac{5\times 10^{-5}}{H_0} \ .
\end{aligned}
\tag{224}
$$

In contrast, the radius of the last scattering surface at $t_{\rm ls}$ is equal to (c.f. Eq. (193))

$$
r_{\rm ls}(t_{\rm ls}) = a(t_{\rm ls})\int_{t_{\rm ls}}^{t_0} \frac{dt}{a(t)} = \frac{a(t_{\rm ls})}{H_0}\int_{a_{\rm ls}}^{1} \frac{da}{\sqrt{\Omega^0_\Lambda a^4 + \Omega^0_m a + \Omega_\gamma(1+r)}} = \frac{3\times 10^{-3}}{H_0} \ . \tag{225}
$$

Clearly $r_{\rm ls}(t_{\rm ls}) \gg d_H(t_{\rm ls})$. What it implies is if you consider two CMB photons arriving at the observer today with coming from opposite directions. They originate from two sides of the last scattering surface, and they have not established causal contact where they were released. This makes it extremely puzzling why both have temperature 2.7 Kelvin.

This is the horizon problem.

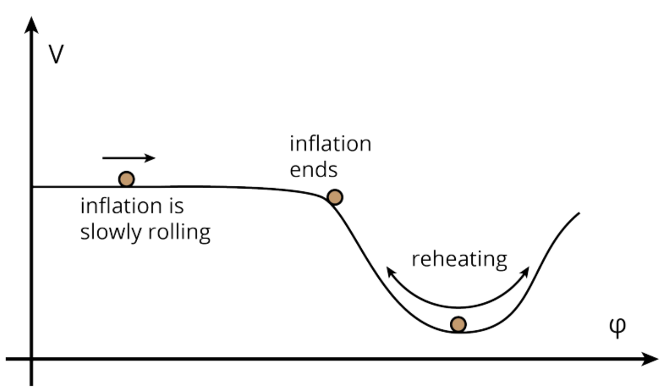
As a simple academic exercise, we could go one step further and ask what is the largest angular separation of two points on the last scattering surface that were causally connected (within particle horizon with each other). This angular size can be computed as

$$\delta\theta = \frac{d_H(t_{\rm ls})}{r_{\rm ls}(t_{\rm ls})} = 0.022 = 1.3^\circ \ . \tag{226}$$

There are also other problems such as the flatness problem, magnetic-monopole problem, etc. We do not elaborate on those in details, because the case is already pretty strong that something beyond the standard big-bang theory is demanded for addressing the horizon problem.

The idea of inflation was suggested by Alan Guth in 1981 (`https://journals.aps.org/prd/abstract/10.1103/PhysRevD.23.347`). It was proposed that at very early time, before the radiation domination era, the universe has undergone a stage of fast expansion. The fast expansion plays the role of separating two points that used to be causally connected to very large distances. The distance has to be larger than the size of particle horizon right after inflation ends and radiation takes over dominating the universe. Through this trick, two points that appear to be causally disconnected (for an observer at later time) actually enjoyed the thermal contact before.

The easiest way to realize the above idea is to postulate a homogeneous inflaton field whose potential energy is much larger than its kinetic energy. This could be the case if the inflaton field is falling down a nearly-flat hill. This is called the slow-roll inflation (see the picture below). Without going to details, I just conclude that this creates a fluid with equation of state $w \simeq -1$. In other worlds, the potential energy of inflaton field plays the role of a "cosmological constant". (Note the value of potential energy needs to be much larger than the cosmological constant in our universe today, otherwise it cannot beat the energy density of radiation at very early times.) Inflation ends when the slow-roll breaks down and inflaton eventually decays into radiation species that fills up the universe.



Based on the discussion in Section 11, we understood that during inflation, the Hubble parameter is a constant H_I, and the scale factor grows exponentially with time

$$a(t) = a(0)e^{H_I t} \ . \tag{227}$$

Before, we have assumed the scale factor is zero at the beginning of universe, which corresponds to $t = -\infty$. In this case, the particle horizon during inflation is infinitely large. Any two points in the universe are causally connected. This is sufficient to address the horizon problem, although choosing the beginning of time to be $t = -\infty$ is not necessary.

Alternatively, we could simply choose the very beginning of time to be $t_i = 0$, with a scale factor $a(0) \neq 0$. In this case, inflation did not happen for infinitely long period of time, and theoretically it is more realistic to build a model (slow roll potential) for inflaton that ends within finite time. The key point is, if inflation lasts for sufficiently long, the exponential factor $e^{H_I t}$ can be made very large. A sufficiently large exponential factor can already provide the necessary ingredient for solving the horizon problem. Suppose inflation ends at time $t_{\rm end}$.

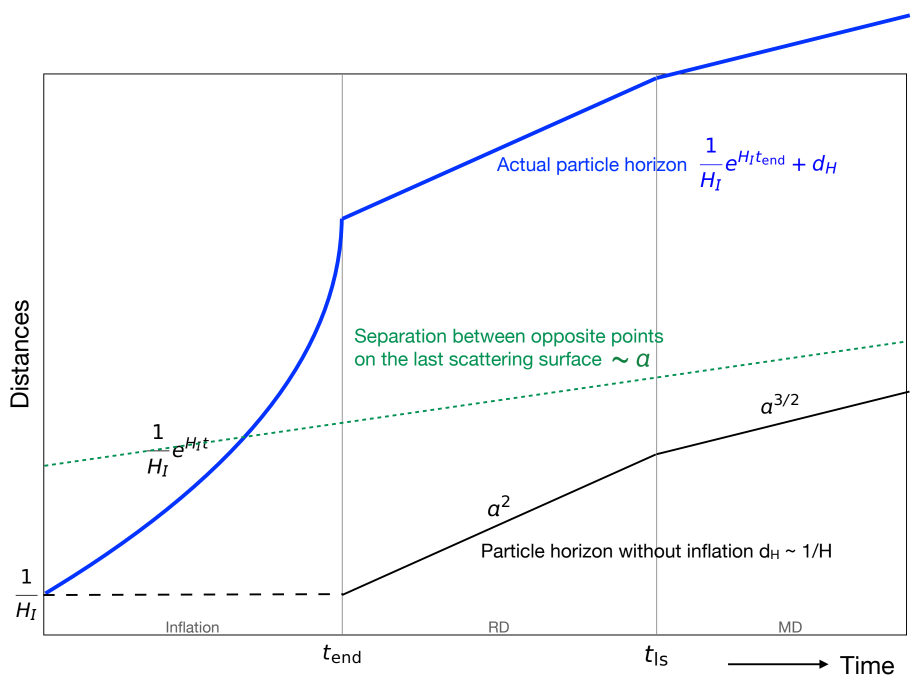
The particle horizon at the end of inflation is

$$d_H(t_{\rm end}) = a(t_{\rm end}) \int_{t_i}^{t_{\rm end}} \frac{dt'}{a(t')} = \frac{a(t_{\rm end})}{a(0)} \int_0^{t_{\rm end}} e^{-H_I t'} dt' = \frac{1}{H_I} e^{H_I t_{\rm end}} = \frac{1}{H_I} e^N \ . \tag{228}$$

In the last step we define the number of e-folds of the inflation as

$$N = H_I t_{\rm end} = \log \left(\frac{a_{\rm end}}{a(0)} \right) \ . \tag{229}$$

As a result, the actual particle horizon at the moment of CMB photon last scattering is the sum of $\frac{1}{H_I} e^N$, and d_H evaluated without including inflation (only RD and MD epochs). The sum can be much larger than $r_{\rm ls}(t_{\rm ls})$, thanks to the exponential expansion. A schematic picture of how inflation solve the horizon problem is shown below. Clearly we need $N \gg 1$. Our remaining exercise is to quantify how large N needs to be.



On the last scattering surface, the largest distance between two points is $\ell_{AB,\text{physical}}(t_{\text{ls}}) = 2r_{\text{ls}}(t_{\text{ls}}) = 5.6 \times 10^{-3}/H_0$. This value corresponds to time of last scattering where the scale factor is a_{ls}. Going back in time to the end of inflation, the corresponding distance was

$$\frac{\ell_{AB,\text{physical}}(t_{\text{end}})}{\ell_{AB,\text{physical}}(t_{\text{ls}})} = \frac{a_{\text{end}}}{a_{\text{ls}}} \simeq \frac{T_{\text{ls}}}{T_{\text{end}}} \; . \tag{230}$$

Here we assume that after inflation the universe is immediately radiation dominated, with a temperature T_{end}. The temperature at last scattering is $\sim 3000\,\text{K}$.

For inflation to successfully solve the horizon problem, we need $\ell_{AB,\text{physical}}(t_{\text{end}}) < d_H(t_{\text{end}})$ so that the two points A and B can be in causal contact. This leads to

$$e^N > \frac{5.6 \times 10^{-3} T_{\text{ls}}}{H_0} \frac{H_I}{T_{\text{end}}} \simeq 10^{30} \frac{H_I}{T_{\text{end}}} \quad \Rightarrow \quad N > 69 + \log \frac{H_I}{T_{\text{end}}} \; . \tag{231}$$

In the above discussion, the inflaton was treated as a classical field. Quantum mechanical effects do exist and leads to other important consequences of inflation:

- Quantum fluctuations in the inflaton and metric fields provide seed for the inhomogeneous universe.
- Quantum fluctuations in the metric tensor also produce gravitational waves, which can lead their imprints in CMB polarization observables.

For a good review of inflation and further reading, see lecture notes by A. Riotto, https://arxiv.org/pdf/hep-ph/0210162.pdf.