Lesson 7 : Temperature history of the universe

Notes from Prof. Susskind video lectures publicly available on YouTube

Introduction

Before talking about temperature, let's go back briefly over what we said at the end of last chapter.

The best way to go about trying to estimate the rough parameters of the expanding universe is to begin with a model, that is Friedmann equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{c_R}{a^4} + \frac{c_M}{a^3} + \Lambda - \frac{k}{a^2} \tag{1}$$

with a set of numerical values for the four parameters c_R, c_M, Λ and k. This is a differential equation in $a(t)$. We solve it for a. That give us a history of the scale factor.

From our model we calculate some of its theoretical consequences, namely the apparent luminosity $L(Z)$ of a candle as a function of the redshift Z, and the density of candles dN/dZ as a function of Z, and we compare them with actual data from observations.

If that doesn't fit so well, we adjust the parameters of the model. And we iterate the process of adjustment until predictions satisfactorily agree with observed data.

We can also do it the other way : we can take the data and try to run the model backwards using the same equations for the luminosity and the density of candles. But it is often harder to do.

Since the specification of the four parameters in equation (1), or that of $a(t)$ and k, are equivalent, we can say that the model that we put in consists basically of two things :

a time history for the scale factor $a(t)$, and a specification of whether we are talking about a negatively curved, positively curved, or flat universe.

Until we get to the properties of the cosmic microwave background (CMB) and some of the detail that is contained in the CMB – which we are coming to in this lesson – we typically don't really see far enough in the universe, or equivalently deep enough in its past, that the curvature parameter k is very critical to the analysis [1].

It is like saying, if we are on the surface of the Earth and we look out a thousand miles : well the curvature counts, but it is not that strong an effect. If everything else we are doing is not too precise, it is hard to tell whether the surface of the Earth is curved on a scale of a thousand miles. Of course if we work in high precision then we can tell. But if all we are doing is looking relatively nearby, the curvarture does not matter too much.

The irrelevance of curvature at short distances, is true whether the space has positive curvature or negative curvature. In other words, as we have now often explained, locally any smooth space is approximately flat.

So counting galaxies, or supernovae and other candles of that sort, at not such large redshifts, i.e. at redshifts that never get bigger than 4 or 5 or 6 or 7 or something like that, our observations are not terribly sensitive to the curvature.

1. Notice that the curvature parameter k is not directly the curvature of the space. Its curvature is $k/a(t)$. It changes with time. In two dimensions for instance, we are familiar with the fact that a big 2-sphere is less curved than a small 2-sphere.

That is both good and bad. It is bad because it doesn't allow us to deduce from the counting of supernovae and so forth with good precision what the curvature is, whether it is 0, -1, or $+1$. On the other hand, we don't need to know too well what the curvature is to draw interesting conclusions about other things. The other important thing, beside curvature, is the history of the scale factor, that is how it varies with time.

So we start with a model, a scale factor that is a function of time, and we calculate two things. One is the intensity as a function of the redshift, in other words what is the brightness of standard sources, standard candles, as a function of the redshift. That is roughly speaking calculating the Hubble law, which is the relation between distance and velocity.

The second function, that we calculate from our model, is the number of objects that we see in the sky of any given redshift, in other words the density of things out there as a function of redshift.

Those are two interesting things which we can calculate from the model. We compare with observations. And that way we pin down the parameters as best we can.

Just from the counting, we don't do as well as if we knew the curvature with great precision – which take CMB data to obtain.

But still, from this process of fitting model and observation, the tendency is toward a fairly flat universe, with no indication of either positive or negative curvature, and toward an expansion history which is consistent with the existence of a large cosmological constant. The cosmological constant corresponds to 70% dark energy today. Said another way, 70% of the energy in the universe today appears to be dark energy.

So that's what we learn from the counting of supernovae or standard candles.

It takes the other end of things, looking very deep, much deeper than where there are galaxies, or much deeper than where there are supernovae or other recognisable standard candles, looking very deep into the microwave background to do better.

So today we start to talk about the microwave background. More generally what we want to talk about is the temperature history of the universe.

Temperature history of the universe

First of all what is temperature ?

We all know what is temperature. We put our hand on a hot flame, it burns. We put it inside the freezer compartment of the fridge, it is painful too.

Strictly speaking, however, temperature is a feature of thermal equilibrium. Thermal equilibrium is a condition for a system that is established after a period of time, basically by scattering.

The kind of systems we are thinking about are systems made up out of particles. What particles ? Well, the particles that are present in the universe. There are photons, there are electrons, there are protons, and nuclei. We forget free neutrons because they decay very quickly [2]. But there are nuclei.

But for our simple purposes let's just say there are electrons, protons and photons. Those are the particles that make up the universe today, at present times.

And thermal equilibrium is what happens after we put particles together and allow them to collide with each other for a relatively long period of time. How long ? Long by comparison with the time scales that it takes for actual collisions to take place. Particles collide with each other in microscopic amounts of time. There are mean free paths between particles.

So there are various timescales. We are not talking about those associated typically with the grand cosmological features of the universe, which are expressed in thousands or billions of years. They are just associated with the nature of the gas or fluid or whatever it is we are talking about. The timescales have to do with the microscopic and macroscopic aspects of the fluid : the timescale of collisions, and the timescale for the fluid to come to rest.

2. Free neutrons have a mean lifetime of about 15 minutes before they decay. They can decay into one proton + one electron + one electron antineutrino.

The fluid might be here in the laboratory. When we say fluid, we mean whatever it happens to be. We just use the term fluid, because it can flow.

When the fluid sits around, if left to itself for a period of time it will establish thermal equilibrium. Even if the gas happens to be in expansion, for example in a chamber whose volume it let to increase, it will still establish thermal equilibrium as long as the expansion is much slower than the microscopic timescale over which these collisions are taking place [3].

But we have got to be careful about the idea of thermal equilibrium. What causes things to become thermalized, i.e. to become in thermal equilibrium ? Let's take again, as our model, our mythical box, figure 1.

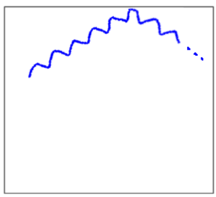

Figure 1 : Box with photons inside. Shown is one photon.

We give the box perfectly rigid and reflecting walls as usual. So we don't have to worry about heating the walls, or energy

3. On the other hand, considering a gas inside a chamber with a piston, if we could pull the piston extremely fast, we would have for a brief period of time all the gas in one part of the chamber. That would not be a slow expansion. And during this period, there would be no meaningful temperature inside the chamber.

being absorbed one way or another by the walls.

We inject photons into the box. For instance, through a little hole in a wall, we shine a laser. Radiation gets inside, and we close the hole. Then the radiation bounces off the walls of the box.

If this is just pure photons, then the box *will not come to thermal equilibrium*. The interaction between photons is indeed extremely weak. Under ordinary conditions, they do not readily scatter off each other.

All that will happen is the pulsive radiation will bounce back and forth and back and forth forever and ever – almost forever and ever – without anything interesting happening, without the energy being spread through the system, without the energy becoming thermal.

What it takes for thermal equilibrium to happen is some scattering. It takes some form of collisions which scatters the photons to cause the system to come to thermal equilibrium.

Photons do scatter a little bit off each other. There is some cross-section for the interaction of photons. But for photons of wavelength comparable to what we have in the universe today, the cross-section for scattering, for the probability for two photons to scatter, is so negligible that the photons inside the box will not equilibrate. They will not come to thermal equilibrium.

What it take to make photons to come to thermal equilibrium is some charged particles. Charged particles scatter off radiation very efficiently.

Consider an electron, and a photon or an electromagnetic wave going past it. The electric field causes the electron to vibrate. The vibrating electron then sends off more radiation, etc. The resulting scattering is very efficient.

If the box contains some density of electrons, figure 2, then scattering will be efficient.

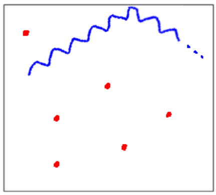

Figure 2 : Box with photons and electrons (red dots e^-).

Let's denote a generic electron by e^-. The minus sign is because electrons have a negative charge[4]. So if the box has some population of electrons in it, then the box will very quickly come to thermal equilibrium.

---

4. It is a *convention* that goes back to Benjamin Franklin. In the XVIIIth century, experimenters had noticed that there appeared to be two kinds of electricity, that created by rubbing a glass stick with silk, and that created by rubbing an amber stick with fur. Franklin decided to attribute a + sign to the former and a − sign to the latter. It can be viewed as an unhappy choice, because an ordinary electric current going down from a higher potential point to a lower potential point is actually a flow of electrons going up the other way.

Of course it depends on what the distances between the electrons are. But if we put a reasonable density of electrons inside, then photons scatter, electrons scatter. Photons and electrons eventually all form a thermal soup. And under those circumstances we would say that *the box has a temperature.*

Now the universe appears to be electrically neutral. There are good reasons for it, which we won't study in this lesson. Suffices to consider it neutral. There even are no large lumps of positive charge, or no large lumps of negative charge in the universe. And certainly on the average it appears to be electrically neutral.

So it is not just photons and electrons, like in the box in figure 2. In addition to electrons, there are protons and other atomic nuclei with positive charge.

Let's enrich our model box and, keeping it simple, just say that there are photons (γ), electrons (e^-) and protons (p). γ ordinarily stands for gamma ray, but the photons are not necessarily only gamma rays, which are very high energy photons. Here we use γ to mean simply photon. So that is what is in the box : gammas, electrons and protons.

Consider the number of electrons, N_{e^-}. It is the same as the number of protons, N_p.

$$N_{e^-} = N_p \tag{2}$$

That is why it is electrically neutral. Well, that is not *why* it is electrically neutral. The *why* of why the universe is electrically neutral is more complicated. It simply states that the model box is electrically neutral.

Of course we use the fact that the electric charge of the electron is equal and opposite to the charge of a proton. That is another experimental fact that we won't try to explain in this lesson.

There are two possible situations in the box :

a) electrons and protons are unbound, freely moving with respect to each other, or
b) electrons and protons are tied together into hydrogen atoms.

Today, in the universe, we are in the *second situation*. Electrons and protons are tied together, forming hydrogen atoms [5]. And these particles are electrically neutral.

Electrically neutral objects are not efficient scatterers. It is true that they do scatter radiation a little bit. But on the whole they are very weak scatterers.

So the very small, dilute density of hydrogen atoms in the universe today has so little effect on the photons moving through it that it would take forever – meaning a very long times, much longer than the age of the universe – to cause the photons to come to thermal equilibrium.

Therefore, *today, the universe is not in thermal equilibrium. It doesn't have a proper temperature.*

5. Hydrogen, in its monoatomic form, and whose nucleus is made of simply one proton, is the most important element in the universe today. It makes about 75% of all baryonic mass.

Another way of saying it is that the various kinds of particles in the universe today are not in equilibrium between each other.

In fact suppose we were to characterize the hydrogen atoms that are out there in the universe by a temperature.

How do we characterize a gas of particles by a temperature ? Roughly speaking the temperature of a gas is the kinetic energy of its particles.

So if we were to assign one temperature to the dilute gas of monoatomic hydrogen in the universe, and suppose we could assign another to the photons, then the temperature of the atoms of hydrogen would be much lower than the temperature of the photons out there.

So the ordinary massive particles, electrons and protons, bound together into hydrogen atoms, are not in equilibrium with the photons in the universe. There is simply not enough scattering for that to happen.

On the other hand, if the electrons and protons were not bound into atoms but freely moving – that is if we were in the *first situation* above – then they would be efficient scatterers. They would come to equilibrium together with the photons, and we would expect them all, γ, e^- and p, to be characterized by a single temperature.

Let's emphasize it : in the first situation, the temperature of the electrons, the temperature of the protons, and the temperature of the photons will all be the same temperature. So, in that circumstance, after thermal equilibrium has been reached, there is a well defined notion of temperature in the box.

What does it take to make sure that the electrons don't fall in with the protons and form atoms which cease to be scatterers ? It takes a high enough temperature.

When the temperature is high enough, the energies of the photons, the energies of electrons, the energies of the protons and so forth, are large enough that enough collisions take place that keep busting up the hydrogen atoms that are formed.

There are some hydrogen atoms when the temperature is high. But there are also plenty of free electrons and free protons. Not only do they keep busting the atoms, thereby creating an equilibrium, but they scatter radiation. That maintains the thermal equilibrium.

So if we are in the regime of temperatures where a temperature is really meaningful, in other words where things are in thermal equilibrium, then the box of gas is described by a number of electrons and a number of protons, or better yet by a density of electrons and a density of protons. And these densities are equal. That is a slightly different interpretation of equation (2) stating that $N_{e^-} = N_p$. And there is a temperature T.

Furthermore, every component of the radiation in the box, every photon if you like, or every bit of radiation, is also characterised by a wavelength λ. There are many different wavelengths in the box once it is in thermal equilibrium. And the distribution of wavelengths is determined by the thermal properties of the fluid in the box.

We can also characterize a radiation, not by its wavelength λ, but equivalently by its frequency ν. The relationship between wavelength and frequency is that the product of the wavelength by the frequency is equal to the speed of light.

$$\lambda \ \nu = c \tag{3}$$

So we can use either wavelength or frequency. And, as we said, there are many many different frequencies. There is a whole spectrum of frequencies inside the box.

The spectrum of frequencies is characterized by a function called the intensity and denoted I. It is basically an energy density. It is a function of both the temperature and either the wavelength or the frequency. So its full notation is for instance

$$I(T, \ \nu) \tag{4}$$

It is defined in such a way that if we multiply it by a small differential volume dV in the box and a small differential frequency $d\nu$,

$$I(T, \ \nu) \ dV \ d\nu \tag{5}$$

that tells us how much energy is stored, in the box, in the differential volume dV, in a band of frequencies between frequency ν and frequency $\nu + d\nu$.

That is the meaning of the intensity. In other words, $I(T, \ \nu)$ is the energy per unit volume and per unit frequency. And it is a function of the temperature – the higher the temperature, the higher the intensity.

We are not going to work through the theory of this intensity function. It is a very important function associated with the radiation of the blackbody, the study of which lead to the discovery of quanta by Max Planck [6] in 1900, thereby revolutionizing physics.

But we can actually work out the behaviour of the intensity function as it would have been worked out before 1900.

Blackbody radiation : ultraviolet catastrophe and Planck's constant

Before 1900, the only constant of nature which appeared in the theory of light and in the theory of radiation was the speed of light c.

Here is a remarkably powerful technique in physics, that we already met a few times in the previous volumes of the collection *The Theoretical Minimum* : we can use simple *dimensional analysis* to ask what must be the formula for the intensity I as a function of T and ν.

Remember that dimensional analysis reasons on the dimensions – in the sense of *units* – of the various quantities appearing in an equation. And it identifies constraints, which are sometimes sufficient to figure out what the equation

6. Max Planck (1858 - 1947), German theoretical physicist.

must be.

The first question is : what is the dimension of the intensity I ? It is energy per unit volume per unit frequency. What is the unit for frequency incidentally ? Inverse time. So whatever the intensity I is, it has units of energy per unit volume, times time. With the conventional notation for units [7] we can write

$$[I] = \left[\frac{E\ t}{V}\right] \tag{6}$$

The other question is : what are the units of temperature ? The units of temperature are basically units of energy.

There is a historical glitch in the definition of temperature. It was not originally defined as an energy. It wasn't even known to be an energy. Temperature was defined in terms of boiling water and freezing water and dividing the gap into a hundred little segments called degrees centigrade or degrees Celsius [8].

But the modern theory of temperature defines it basically as being the average energy of a molecule or particle in a gas, when the gas is in thermal equilibrium – the particles then have a certain *distribution* of energy.

The numerical conversion factor between energy units and conventional temperature units like degrees Celsius or de-

7. See, for instance, volume 1 chapter 3 of the collection *The Theoretical Minimum.*

8. Anders Celsius (1701 - 1744), Swedish astronomer, physicist and mathematician.

grees Kelvin [9] is a certain constant called Boltzmann's constant [10]. So you will often see formulas in which some energy or another E – whatever it happens to be – is written as being proportional to some temperature T times a constant k_B (read k Boltzmann)

$$E = T\ k_B \tag{7}$$

where T is expressed in degrees Kelvin, that is such that the 0 mark corresponds not to freezing water but to motionless particles. We could also use degrees Celsius, but then we would have to add the additive constant 273.15 x k_B on the right-hand side of equation (7).

So the Boltzmann constant is just a conversion factor. We will keep it around because it is traditional. But for our real purposes, temperature is an energy and it has units of energy, k_B being just some historical relic. In fact we could easily set it equal to 1 with appropriate units, and not lose anything, like we often do with c. However we will keep it around because it is customary to keep it around, and to remind the reader that the temperature is really an energy.

Now we can ask : what kind of dependence can the intensity I, which has certain units, have on temperature and frequency, *if it is expressed only with* T*,* ν *and* c?

Frequency has units of one over time. Temperature has units of energy. The speed of light has units of length per

9. William Thomson, aka Lord Kelvin, (1824 - 1907), Irish-born British mathematical physicist and engineer.

10. Ludwig Boltzmann (1844 - 1906), Austrian physicist, father of statistical mechanics.

time. And I has to have units of energy times time, over volume, which is length cubed. We want to express I in terms of T, ν, and c. The reader will readily check that the general formula cannot but be

$$I \sim \frac{T\nu^2}{c^3} \tag{8}$$

Exercise 1 : Check that, with the temperature T, the frequency ν and the speed of light c, the only way to construct a quantity I with units of energy x time over length cubed, is to write

$$I \textbf{ is proportional to } \frac{T\nu^2}{c^3}$$

This is the only formula which has the right units. Of course there can be a dimensionless multiplicative constant in front of the right-hand side of equation (8). Also whenever there is T in one of our formulas, we multiply it by the Boltzmann constant. It is a single entity $k_B T$ which has units of energy.

In fact, pursuing the argument, the exact formula for I turns out to be

$$I = \frac{8\pi\nu^2}{c^3} k_B T \tag{9}$$

This formula is called the *Rayleigh-Jeans law*[11]. It is the

11. named after the two British physicists John William Strutt, aka Lord Rayleigh, (1842 - 1919), and James Jeans (1877 - 1946).

prediction, derived from classical physics, of what the radiation energy per unit volume and unit frequency should be in the box in figure 2.

We write "should be" because actually this formula is a disaster! It was already known to people in the late XIXth century, in the 1890's, that it was a disaster. Here is why.

It we plot I as a function of ν, the Rayleigh-Jeans formula says that it grows like a parabola, figure 3.

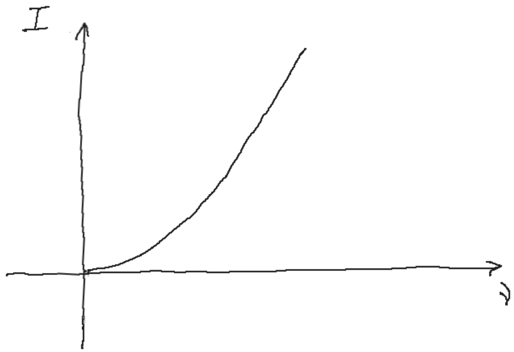

Figure 3 : *Theoretical* density of energy I in the box, per unit volume and per unit frequency, as a function of the frequency ν, according to the Rayleigh-Jeans law.

We readily see that the total amount of energy in the box is then infinite. There is just more and more energy per unit volume and per unit frequency as the frequency goes higher and higher. But that is not possible.

This problem is called the *ultraviolet catastrophe* – ultraviolet because that is the range of frequencies beyond the visible spectrum which ends up with violet.

Equation (9) and figure 3 are what classical physics predicts. But of course experimentally the curve was quite different. After an increase, it turns over and goes back down to zero, figure 4.

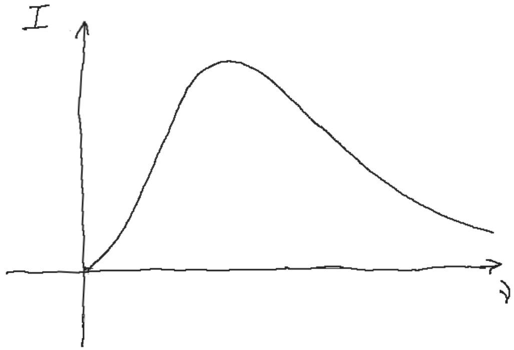

Figure 4 : *Experimental* density of energy I in the box, per unit volume and per unit frequency, as a function of the frequency ν.

We won't go into the whole story of how physicists groped for the solution to the ultraviolet catastrophe. The Rayleigh-Jeans formula fits the experimental curve correctly only for very low frequencies, or equivalently very long wavelengths. Wien [12] had another ad hoc formula which fitted well the experimental curve for high frequencies, that is for the tail part of the curve to the right in figure 4.

To learn the complete story of this most fascinating physical riddle [13], to learn which experiments were conducted to

12. Wilhelm Wien (1864 - 1928), German physicist.

13. Lork Kelvin had famously said in 1900 that physics was essentially complete. Everything, according to him, was by then understood, except for two little clouds. One cloud was the failure to detect

try and measure the intensity as a function of wavelength, the role of guess work, and how it was solved, we invite the reader to turn to specialized books on the history of physics.

So the real curve of I as a function of ν, measured experimentally, is shown in figure 4. After growing, it reaches a peak, then bends over and goes back down to zero. Thus the total energy in the box can remain bounded.

Planck figured out that it was possible to keep the general shape of formula (9), while fitting the experimental curve, if he introduced *a new constant* in it. He proposed the following formula

$$I = \frac{8\pi\nu^2}{c^3}\frac{h\nu}{e^{\frac{h\nu}{kT}} - 1} \tag{10}$$

where h is Planck constant. Equation (10) was still a more or less ad hoc modification of the Rayleigh-Jeans equation by Planck, although he reached it via a hypothetical quantization of energy in the box, see below. And this eventually launched quantum mechanics.

What does formula (10) have to do with formula (9) ? The answer is that the two formulas are the same for low frequencies. Indeed, when ν is small, $h\nu/kT$ is small. Then we apply the familiar approximation for the exponential of a small quantity x

$$e^x \approx 1 + x \tag{11}$$

---

the movement of the Earth through the ether. The other was the ultraviolet catastrophe. Both clouds turned out to be the starting point of tremendous upheavals in physics in the XXth century.

Mathematically speaking, the right-hand side is the beginning of the Taylor-MacLaurin expansion of e^x. The subsequent terms in the expansion are of the form $x^n/n!$ where $n \geq 2$ and can altogether be neglected when x is small.

So, when ν is sufficiently small, we can write

$$e^{\frac{h\nu}{kT}} - 1 \approx \frac{h\nu}{k_B T}$$

Planck's formula (10) can be rewritten

$$I = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{\frac{h\nu}{kT}}$$

or equivalently

$$I = \frac{8\pi\nu^2}{c^3} k_B T$$

which is the same as formula (9).

In other words, for small ν, the formula as written at the end of the XIX$^{\text{th}}$ century by the classical physicists Rayleigh and Jeans, and also by other people, is correct.

But of course, in Planck's formula (10), as ν gets large, the term $h\nu/(e^{\frac{h\nu}{kT}} - 1)$ is no longer approximately equal to $k_B T$. The exponential starts to get very big.

In particular, when $h\nu/k_B T$ becomes bigger than 1, the exponential becomes large. In fact it is quickly so large that the -1 next to it is not even important anymore. The whole factor $h\nu/(e^{\frac{h\nu}{kT}} - 1)$ is equivalent to $h\nu e^{-\frac{h\nu}{kT}}$, that is decreases exponentially.

So there is a crossover point. And it is when $h\nu/k_B T$ is about equal to 1. If we look at the plots again, the curve of I instead of growing forever like a parabola, as shown in figure 3, rather suddenly turns over and goes back exponentially to zero, as shown in figure 4, thus solving the ultraviolet catastrophe.

This observation provides another interesting connection. Let's remember that for any radiation the frequency ν and the wavelength λ are related by

$$\lambda\nu = c \tag{12}$$

So, dropping for lighter notation the index $_B$ in k_B, the inequality

$$\frac{h\nu}{kT} > 1 \tag{13}$$

can be rewritten

$$\lambda < \frac{hc}{kT} \tag{14}$$

So the energy density at high frequency, or small wavelength, is strongly affected by quantum mechanics. The spectrum of energy density, plotted in figure 4, is driven very quickly to zero when ν gets high.

The crossover point it shown in figure 5. It is at

$$\lambda = \frac{hc}{kT} \tag{15}$$

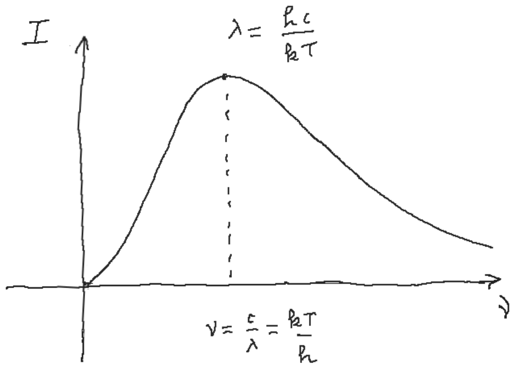

Figure 5 : Crossover point in the spectrum of energy density.

And, more than that, it is where most of the power is – power being used casually in the sense of energy density. The largest part of it is around $\lambda = hc/kT$.

In other words, most of the energy density is in wavelengths that are proportional to one over the temperature, the proportionality factor being the constant hc/k_B, that is Planck's constant times the speed of light divided by Boltzmann's constant.

That allows us to establish a *connection between* λ *and* T. We can sort of use that connection between lambda and temperature even when the system is not in strict thermal equilibrium, just to get a rough sense sometimes of what is going on inside the box[14].

If we know we have a bunch of photons, whose average wavelength is some particular wavelength, and we want to know how potent they are at doing various things to a system, we can roughly speaking replace it by the question of how potent would be radiation at a corresponding temperature.

This link between temperature and wavelength will lead us to the cosmic microwave background (CMB) radiation.

But let's first of all do a questions / answers session.

Questions / answers session

Question : When Planck worked on his formula, was he working with a model of electrons, protons and photons in the box, as shown in figure 2 ?

Answer : No. In the late 1890's, Planck was doing a combination of two things. He was doing curve fitting. And he was searching for a fundamental explanation of the ultraviolet catastrophe within the framework of classical physics[15].

---

14. Also called the *blackbody*, because some variants of the experiment involve a body which doesn't reflect light, therefore is black.

15. Planck, who was 32 years old in 1890, was for a long time a disbeliever in the atomic hypothesis. He was convinced that eventually the view that matter was continuous, infinitely divisible, and so forth, would win over the highfalutin idea of existence of atoms combining into very small moles, called molecules, to explain

Various things were known. The fact that the peak wavelength, shown in figure 5, shifts linearly with temperature was known. It was called *Wein displacement law.* Plank knew that very well. He expressed the formula for the displacement law linking the peak wavelength and the temperature with the help of a new constant.

The Boltzmann constant k_B was nothing new. Everybody knew, by that time, that k_B and temperature went together. And the speed of light was known.

So Planck introduced the constant h to express Wien displacement law as

$$\lambda_{peak} = \frac{hc}{kT} \tag{16}$$

And he gave the numerical value of h.

---

various puzzling observations made by chemists and others.

For instance, chemists had observed that one liter of oxygen combines only with 2 liters of hydrogen to make water. If we try to put more hydrogen we are left we some hydrogen, and if we try to put more oxygen we are left with some oxygen. The atomic hypothesis, plus some other hypotheses pertaining to thermodynamics and chemistry, explained that nicely.

In the face of mounting evidence for the existence of atoms, however, in the 1890's Planck finally accepted the atomic hypothesis. He became very familiar with the works of Boltzmann, who had recast thermodynamics as a statistical theory about the behaviour of huge numbers of particles forming a gas.

But Planck naturally remained, as everyone else at the time in physics, a classical physicist.

Moreover he came up with his formula (10), also involving h. It was essentially a curve fitting to be in accordance with the experimental curve of $I(T,\ \nu)$ shown in figure 5.

It can be said that the most important tangible part of what Plank did was to identify this constant h. But the way he was lead to it was much deeper than it looks.

Indeed, less tangible but more important, Planck's contribution which eventually lead to a revolution in physics was his speculation that the behavior of the whole blackbody system might have something to do with harmonic oscillators [16]. Without entering into his calculations, that is in fact what lead him to the constant h, which appears in formula (16) for λ_{peak} expressed as a function of K_B, T, h and c, and in formula (10).

His intuition was brilliant. But then he searched in the wrong direction. He thought that the walls of the box contained vibrating things of some sort. It didn't occur to him that it could be the radiation itself that had to do with harmonic oscillators, that it could be described with oscillators, because that could not make sense in the framework of classical physics.

Remember that we are before atoms were known for sure. Molecules were not really understood. So even these mechanical vibrations from the walls were a bold hypothesis.

16. He said later that it was a last and desperate hypothesis to try to fit the curve of I with a formula valid for all ν's, and whose integral over ν from 0 to $+\infty$ converged, thereby solving the ultraviolet catastrophe, which was his ultimate objective.

So, like every great physicist, Planck had some right ideas and some wrong ideas.

It was Einstein who derived Planck's formula (10) in detail from a set of principles. But that is not where we want to go in this lesson.

Q. : What are the units of Planck's constant ?

A. : It can be calculated from the following observation : in Planck's formula (10), the denominator $e^{\frac{h\nu}{kT}} - 1$ is dimensionless because 1 is dimensionless. Hence $h\nu/kT$ is dimensionless too.

Therefore h has the same dimension as kT/ν. Remember that kT is an energy, and ν is the inverse of a time. So h has units of energy times time.

It can also be obtained from the well-known formula $E = h\nu$ for the energy of a photon of frequency ν.

Cosmic microwave background radiation

Replacing our blackbody by the universe, we can look at the photons that are in it today, bathing somehow the great intergalactical emptiness. The overwhelming majority of the photons that we see in the universe today have a particular wavelength.

The wavelength is about a millimeter. And that allows us to associate with it a temperature. That temperature is about three degrees Kelvin.

Let's say right away that we should not interpret this fact, in itself, as saying that anything is in thermal equilibrium. It is just a rough correspondence.

In fact, it will turn out to be much better than just a rough correspondence, for reasons we will come to.

Let's look again at Planck's formula (10), get rid of the bunch of multiplicative constants $8\pi h/c^3$, and simply write it as a proportion

$$I \sim \nu^3 \frac{1}{e^{\frac{h\nu}{kT}} - 1} \tag{17}$$

Now let's divide and multiply by the temperature cubed

$$I \sim \frac{\nu^3}{T^3} \, \frac{1}{e^{\frac{h\nu}{kT}} - 1} \, T^3 \tag{18}$$

What we are interested in is the *shape of the curve* in figure 5, not the height of the curve. So it is the same as the shape of

$$\frac{\nu^3}{T^3} \, \frac{1}{e^{\frac{h\nu}{kT}} - 1}$$

as a function of ν.

And it is a function of ν only through the ratio ν/T. In other words, it has a universal shape

$$F\left(\frac{\nu}{T}\right)$$

such that if we change the temperature, all that happens is that the curve rescales along the frequency axis.

For example supposing we multiply the temperature by a factor of 2, then the curve simply rescales along the ν axis. It stretches by a factor 2. If we want to change the temperature by a factor of 100, we just stretch the curve by a factor of 100.

In terms of wavelength, if we remember that ν is proportional to one over the wavelength, a similar thing happens. But now the rescaling is not a stretching but a squeezing. If we multiplied the temperature by 2, we would squeeze by a factor 2 the shape of the curve that we would have if we used λ instead of ν on the horizontal axis. In particular, the peak wavelength is divided by 2.

In summary, *there is a universal shape of the blackbody spectrum.*

If you know it at one temperature and you want to get it at another temperature, you simply rescale it. You rescale the frequency proportionally to the rescaling of the temperature. That is an important point that is going to come up. And it is a simplifying fact.

In the universe today there are many more photons than there are protons and electrons.

If these protons and electrons were free they would be scatterers for the photons. It would lead to thermal equilibrium.

But most of the protons and electrons are bound up into atoms. Moreover they are very dilute. There are not very many of them per unit volume.

As we said, atoms are neutral, hence not efficient scatterers. Therefore there is not much possibility for the photons to be scattered. So the universe today is not in a process of thermal equilibration, and in fact is not in thermal equilibrium.

The fact that there is a temperature associated with the blackbody radiation is a kind of accident. We shall presently explain what this accident is.

When the temperature is too low in the box, the protons and electrons are bound into atoms. Then they don't scatter much the photons. We say in that case that *the radiation has decoupled* from the charged particles, figure 6.

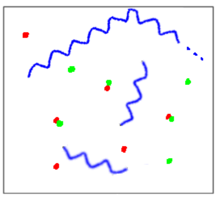

Figure 6 : Box containing photons, protons and electrons. Below a certain temperature, protons and electrons are all bound into atoms, and radiation decouples from the charged particles.

When there is decoupling, for the most part the radiation in the box doesn't even see the charged particles. It is kind of two independent systems in the same space.

Now let's imagine we start with the box at a high temperature. For whatever reason it is hot. Suppose it is so hot that the electrons and protons in it are ionized – that is not bound together to form atoms, but free. Or at least that not all the protons and electrons are bound into atoms.

Then, as we saw, the box is in thermal equilibrium. And it does have a temperature. Furthermore the radiation spectrum looks like the blackbody radiation spectrum.

Now let's start to expand the box. As its volume increases, the box cools. It cools because the various particles do work on the walls of the box. It is the product of the forces from pressure on the walls times displacements. And the work depletes the energy inside the box.

The temperature inside the box goes down of course. Eventually it gets so cold in the box that the electrons and protons combine into atoms. Then there aren't enough high energy photons to re-ionize the atoms. And there aren't enough free charged particles to scatter the photons.

So what happens ? *The radiation decouples from the atoms.* The atoms and the photons each go their own way. All sorts of things may happen to the atoms. Their temperature may continue to go down. But that is not what we are interested in. It is the photons that we are interested in.

The temperature of the radiation goes down too. However the universe as a whole doesn't stay in thermal equilibrium. For a system to be in thermal equilibrium, the temperature of every part of it must be the same [17].

So the system formed by the photons and the atoms in the universe falls out of equilibrium. Nevertheless as we continue to expand our model box of the universe, the shape of the blackbody spectrum, related to the radiation, is sort of frozen in.

As *each linear dimension* of the box is multiplied by a factor a, the wavelength of every single photon in it is simply stretched by the same factor a. As we already saw, the shape of the spectrum doesn't change. Indeed it is a universal shape.

Consequently the central wavelength, in figure 5, also gets multiplied by a. Therefore, according to the formula $\lambda_{peak} = hc/kT$, the temperature T is divided by a. In other words, expanding each linear dimension of the the box factor a mimics letting the radiation inside cool off by the same factor.

Notice that the reasoning is only true for radiation. What happens to the other particles in the box ? They do whatever they do. But once they have coalesced for good into atoms, they don't couple very much to the radiation anymore. And they will follow their own temperature history.

Radiation on the other hand, which had a blackbody spectrum and a well defined temperature before decoupling, will

17. We haven't defined very precisely what it means for a system to be in thermal equilibrium, relying instead on the intuitive notion from elementary physics. A precise definition is provided by the statistical thermodynamics theory of Boltzmann, which is the subject of volume 6 in the collection *The Theoretical Minimum.*

retain its blackbody spectrum after decoupling, except insofar as it will rescale according to the linear expansion of the box. Therefore, even though there are no longer particles to efficiently scatter the photons and maintain their thermal equilibrium, they will retain a temperature. And that temperature will go down in inverse proportion to the expansion of the box.

That is why today, when the universe is far too cold to be in thermal equilibrium, which would imply radiation and matter to be at the same temperature, nevertheless the spectrum of radiation in the universe has the blackbody form.

We emphasize it : it is not because the radiation is in equilibrium with anything, it is just because it had this shape frozen in at some very early time. And it maintained that shape.

Now, how do we know that it had that shape at very early time ? We know it because we measure the shape of the photon spectrum in the universe. There is of course lots of different kinds of photons in the universe. There is gamma rays, there is ordinary starlight, and many other kinds, which don't all date from the time of decoupling.

But the overwhelming majority of photons that we see are in the wavelength range of about a millimeter. Anything else is just more or less incidental, carries altogether very little energy, and is fully accounted for by mechanisms that we fully understand.

The dominant form of radiation that we observe today, both in terms of energy and by counting the number of photons, is, by a long shot, the blackbody cosmic microwave background radiation.

It was discovered in 1964 by Penzias and Wilson [18], and has since been extensively studied. It does have, to a very high precision, the shape of the blackbody spectrum, figure 7.

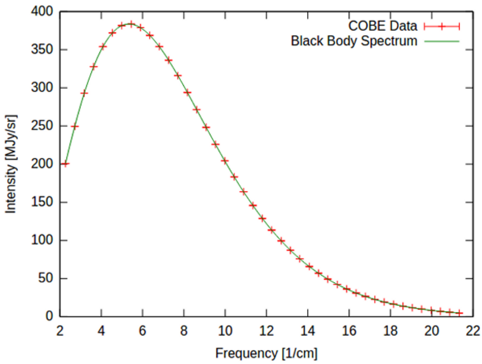

Figure 7 : Cosmic Microwave Background Spectrum from the COBE satellite. The error bars are too tiny to be visible on the graph. Source : NASA.

So what we see today is a universe filled with radiation that appears for all practical purposes to be the remnants of a thermal distribution that must have become themalized, a long time ago, at a much higher temperature when the electrons and protons were ionized. That is of course our best guess as to what the thermal history of the universe is. But let's expand on that a little bit.

18. Arno Penzias (born in 1933), and Robert Wilson (born in 1936), American physicists and radio astronomers.

Questions / answers session (2)

Question : A photon that travels through vacuum, how much space does it occupy ?

Answer : That is a loaded question.

In some sense the photon has no size at all. But it has a quantum mechanical spread. Let's call it a probability distribution.

Let's say it occupies a size roughly equal to its wavelength, that is to the inverse temperature according to the formula $\lambda_{peak} = hc/kT$.

Most of the photons that are floating around not only have a wavelength of about that big, but the wave packets that describe them are also roughly about that big.

So the best answer is a millimeter. The characteristic size scales of the photons, the wave packets, the wavelength, are all the same, about a millimeter in today's universe.

You have to realize that there is no universal, well defined notion of how big a particle is. There are several different things that you could call the size of it. So I'm not going to try to answer that.

Temperature at decoupling

What can we say about the CMB ? Or what does it say to us ? First of all, what are the facts and what do we know ?

There is another formula that we want to write down, and that we will come back to.

The phenomenon where we start with ionised gas, we lower the temperature, and we come out with some atoms is called *recombination.* It is called recombination, but it is simply combination. It is where electrons and protons combine into atoms.

It is the point at which the scattering stops too. And that is called *decoupling.* It is characterized by a certain temperature. And it is characterized by a wavelength through the connection between wavelength and temperature.

What about the wavelength of the radiation today ? Let's write the formula. It is sort of a trivial formula. The wavelength of the radiation we see today, the remnant of the radiation from the blackbody spectrum before decoupling, let's call it λ_{today}. We mean the average wavelength, the wavelength at the top of the peak [19].

We are interested in

$$\frac{\lambda_{today}}{\lambda_{decoupling}} \tag{19}$$

19. We assimilate casually, for the statistical distribution, the maximum likelihood and the average, as is usual in physics, except for very skewed distributions.

Is it a big number, or a small number ? A big number of course. The wave got stretched.

What did it get stretched by ? It got stretched by the expansion of the universe.

So we can write that the ratio (19) is just the ratio of the scale factor today divided by the scale factor at the time when decoupling took place.

$$\frac{\lambda_{today}}{\lambda_{decoupling}} = \frac{a_{today}}{a_{decoupling}} \tag{20}$$

The time at which decoupling took place is not an absolutely rigorously, sharply defined time. It happened over some period of time. But the characteristic time, in the sense of period, over which it happened is relatively short. And so we can talk about the ratio of the scale factors.

In other words, the ratio (19) is also the ratio (20) by which the universe expanded over the period of time between the decoupling phenomenon and today.

So it is interesting to ask : what do we know about it ?

Since the wavelength is inversely proportional to the characteristic temperature, the ratio (19) is also equal to the ratio of the temperatures as follows

$$\frac{\lambda_{today}}{\lambda_{decoupling}} = \frac{a_{today}}{a_{decoupling}} = \frac{T_{decoupling}}{T_{today}} \tag{21}$$

Whose temperature today ? Not the temperature in the room. Not even the temperature we would feel in outer space if we were out there floating around. The temperature we would feel would be mostly the temperature of the dilute molecules and atoms surrounding us. But that is not what we mean by T_{today}.

By T_{today}, we mean the temperature of the cosmic microwave background radiation today. That temperature is about three degrees Kelvin.

So another interesting question is : what can we say about the temperature at decoupling ? What was the temperature at which hydrogen recombined. Or, better yet, if we go upstream in time, what would be the temperature at which hydrogen would ionize ?

There is a one simple answer to it : it is when the temperature is such that the photons have enough energy to ionize the atoms.

The energy of ionization of the hydrogen atom is 13.6 electron-volts. Thus if the characteristic photons, that is those corresponding to λ_{peak}, have an energy of 13.6 eV, then they will have plenty of energy to kick the hell out of the hydrogen atom and ionize it.

That actually corresponds to a much higher temperature than necessary, and therefore much higher than the decoupling temperature $T_{decoupling}$. We will see why in a moment.

Anyway let's understand the schematic reasoning, and then refine it. In a simple theory of the temperature of decoupling, it is the temperature at which the average photon has an energy of 13.6 eV, and therefore will ionize an atom made of one proton and one electron, i.e. a hydrogen atom.

Why does the decoupling occur at a lower temperature than that where the average photon has an energy of 13.6 eV ? That has to do with the statistical distribution of energies of the photons.

There is another element that we have to take into account : there is a lot of photons out there. The ratio of the number of photons N_γ to the number of protons N_p, or the number of electrons N_e, or of atoms – there are all the same number – is very large.

$$\frac{N_\gamma}{N_e} \approx 10^8 \tag{22}$$

Now, where does that number come from ? How did so many photons get there ? It is a question we are going to address.

But the ratio (22) is a fact from observation. We know how many protons there are, or how many electrons there are. We can also measure how many photons there are from the blackbody spectrum. And the ratio of the number of photons to the number of electrons is about 10^8.

Now, consider a soup of photons in thermal equilibrium at temperature T. What is the probability that a photon have a given energy ? The answer is given technically by a *density of probability.* At temperature T, the probability for a photon to have an energy between ϵ and $\epsilon + d\epsilon$ is

$$\frac{1}{kT}\ e^{-\epsilon/kT}\ d\epsilon \tag{23}$$

where ϵ is the variable. The function which multiplies $d\epsilon$ is a standard function in statistical thermodynamics. It is called the *Boltzmann distribution*. It is a density of probability.

We are progressing toward our objective of figuring out the temperature of decoupling.

Now we can ask the question : what is the probability that a photon have the ionization energy, or, more precisely, have the ionization energy or more ? The ionization energy is ϵ_{ion} = 13.6 eV. It becomes a simple probability calculation on the tail of the density given by equation (23). And let's not forget that the density falls off quickly. That allows us to speak casually about the probability of photons having that energy to mean those at that energy and above.

The answer is

$$P\{\text{ energy } \geq \text{ 13.6 eV } \} = \int_{\epsilon_{ion}}^{+\infty} \frac{1}{kT}\ e^{-\epsilon/kT}\ d\epsilon \tag{24}$$

Integrating the density under the integral sign yields

$$P = e^{-\epsilon_{ion}/kT} \tag{25}$$

Now, since there are 10^8 times more photons than protons in the universe, there is another way to formulate the question. Notice by the way that ratio of 10^8, that holds today, also held at decoupling. It is only when the temperature was even much higher and capable of altering the protons themselves that the ratio could be different.

But the history since the decoupling time is pretty much a history of conserved numbers of particles in the universe, conserved number of photons, conserved number of electrons, conserved number of protons. To figure that out requires some calculation, but the botton line is that we can consider that the number of particles of any given species didn't change very much.

So let's reformulate the question as follows : considering one proton, and the 10^8 photons that correspond to it, at which temperature T the probability that there is at least one photon of energy 13.6 eV will be close to one ?

Since 10^8 is approximately e^{20}, a quick and dirty estimation [20] of T is given by

$$e^{-\epsilon_{ion}/kT} \; e^{20} \; = 1 \tag{26}$$

where ϵ_{ion} is the energy at ionization, i.e. 13.6 eV. This implies

$$\frac{\epsilon_{ion}}{kT} = 20 \tag{27}$$

or

$$kT = \frac{\epsilon_{ion}}{20} \tag{28}$$

It means, given that there are so many photons around, that we don't actually have to heat the system up to the

20. It is an elementary probability fact. Suppose that we have a large number, n, of balls, each of which can be red with probability p or not red with probability $(1 - p)$, and p is very small. Then the probability that there is at least one red ball is approximately np, because the probability that there is none is $(1 - p)^n \approx 1 - np$.

ionization temperature before there is a significant population of photons able to ionize the atoms.

At a temperature of only one twentieth the ionization temperature, there are already enough sufficiently energetic photons around to free a substantial proportion of the electrons from the protons.

The probability calculation that we did is very rough in several respects. Its purpose was to illustrate why the temperature of recombination is not that of ionization. The actual temperature of recombination is rather about 1/40 the temperature of ionization.

Questions / answers session (3)

Question : Is the number of photons coming out of the solar system negligible, or are they not high enough energy photons to cause ionization ?

Answer : There are several questions in your question.

First of all, stars, and photons coming out of stars, were not around at the time of decoupling. So we don't care about them at all.

Secondly, the vast majority of photons today in the universe are photons belonging to the CMB. Starlight photons are only a tiny fraction of the population of overall photons.

For these reasons, we are only counting the photons that are in the cosmic microwave background, in other words, the remnants or the relics of that early thermal equilibrium.

Those are the ones that count as far as concerns the elucidation of the history of the universe through the study of the photons today.

Q. : At the time of decoupling, all the electrons and protons combining into hydrogen atoms emit 13.6 eV photons. Should not there be a line then somewhere in the spectrum of the CMB corresponding to that emission ?

A. : As long as the recombination process is happening slowly enough, the photons will stay in thermal equilibrium. And the curve will be the right blackbody spectrum curve.

If the expansion took place a little faster, yes, then there would be a kind of bump at 13.6 electron-volts.

But as long as it happens slowly, the balance will maintain itself. And the population will conform to the thermal equilibrium.

So, technically, if the expansion is slow enough that the process is *adiabatic*, it will remain in thermal equilibrium at all times – until the point where there simply are not enough charged particles to keep it at equilibrium.

Q. : What is that actual temperature of recombination ?

A. : It is about 4000°K. You take the formula

$$kT = \frac{\epsilon_{ion}}{40}$$

and plug-in the values for ϵ_{ion} and k_B.

$$\begin{aligned} k_B &= 1.3806 \text{ x } 10^{-23} J/K \\ \epsilon_{ion} &= 13.6eV \\ 1eV &= 1.602 \text{ x } 10^{-19} J \end{aligned}$$

4000°K is the value we obtain from the ratio $\epsilon_{ion}/40$.

A detailed calculation has to be done to figure out this ratio between the ionization energy and the average energy of the photons at recombination. It has been done. I haven't done it. But this is the basic physics.

The important thing is that it is a good deal lower than the ionization temperature. It is a factor of about 40. So decoupling happens at a temperature of 4000 degrees Kelvin.

Landmarks in the history of the universe

The temperature of the current CMB radiation is three degrees Kelvin.

According to equation (21), the ratio of temperatures between decoupling and today also gives us the ratio of scale factors. 4000° divided by 3° is about 1300. For simplicity, because here we are not concerned with exact figures, let's talk of a factor 1000.

So this is also equal to the scale factor today divided by the scale factor at decoupling.

Thus our study of the recombination process in the past and the CMB today produces the amazing piece of information that the universe expanded by a factor of 1000 from the last time it was opaque till today.

We say *opaque* because before decoupling the particles of matter in the universe were ionized and therefore light could not readily travel through it. So at that time the universe was opaque. After decoupling it became transparent. Now light plays a useful role in letting us see things, whereas before decoupling we could not have seen anything.

Going backwards in time, the decoupling time is one landmark. It is estimated to be 400 000 years *after the beginning of the universe*, which is itself estimated to have taken place 13.8 billion years ago [21].

If we use the scale factor $a(t)$ as a way to chart the history of the universe, that is another way instead of time to label important events, we can say that this ratio of 1000 is one landmark so to speak – meaning that the point where $a(t)$ was 1000 times smaller than today labels the important landmark of decoupling.

21. The date of the beginning of the universe is of course a fascinating date. Rather than talking about the beginning of the universe, some cosmologists prefer to talk about the beginning of its spatial expansion.

Of course there are lots of landmarks, going backwards in time, that we have not mentioned yet in our study of the past history of the universe. We haven't talked about galaxy formation. We haven't talked about black holes forming, etc. We will come back to those things.

We haven't gathered enough information in this course yet to think about how the galaxies formed and so forth. But jumping back and ignoring that, we have met our first landmark which is the decoupling landmark.

The second landmark – going further back in time – we want to look at is when the universe went from being radiation-dominated to being matter-dominated.

Remember what radiation-dominated and matter-dominated mean. The equation of state for radiation and the equation of state for matter lead to two different formulas for the energy density.

In a matter-dominated universe, the energy density ρ_{matter} scales like some constant ρ_M divided by a^3

$$\rho_{matter}(t) \sim \frac{\rho_M}{a(t)^3} \tag{29}$$

That just expresses a dilution of particles as the universe expands.

Whereas in a radiation-dominated universe, the energy density $\rho_{radiation}$ scales like some constant ρ_R divided by a^4

$$\rho_{radiation}(t) \sim \frac{\rho_R}{a(t)^4} \tag{30}$$

As we go forward in time, radiation becomes less important.

Of course it is much less important today. But, as we go backwards in time, we come to a time where formula (30) becomes bigger than formula (29).

Today, and for almost the entire observable period of the universe, it was matter-dominated – meaning that the matter contribution was much bigger than the radiation contribution.

That affected the way the universe expanded. Solving Friedmann equation we saw that during the matter-dominance $a(t)$ expanded like $t^{2/3}$, instead of $t^{1/2}$ that prevailed during the radiation-dominance.

But if we go back early enough, where the scale factor is small enough, we come to the point where there is a crossover between (29) and (30).

Let's see if we can estimate where that crossover happened. What was the scale factor when that happened ?

Let's consider the ratio *today* of the energy density in matter to the energy density in radiation

$$\frac{\rho_{matter}}{\rho_{radiation}} \tag{31}$$

How do we calculate it ?

First of all, we know from the blackbody spectrum how much energy density is present today in the form of photons. And we can use the fact that for every proton, or every hydrogen atom, there are 10^8 photons.

Second of all, we know the energy of each photon. It is about 10^{-4} eV. That is what corresponds to the CMB temperature of 3 degrees.

What about the energy of a proton ? The energy of a proton is 1 billion electron-volts. And let's not forget about dark matter, which is approximately one order of magnitude larger than luminous matter [22]. So we can calculate the ratio (31) :

$$\frac{\rho_{matter}}{\rho_{radiation}} = \frac{10^{-8}\ 10^{9}\ 10\ eV}{10^{-4}\ eV} = 10^{6} \tag{32}$$

The ratio of matter, including dark matter, to radiation today is about 10^6.

The next question is : how did it behave in the past ?

As we extrapolate backward in the past, ρ_{matter} scales with three powers of the scale factor $a(t)$, whereas $\rho_{radiation}$ scales with four factors of the scale factor.

That means the ratio (32) scales with one power of $a(t)$. More explicitely as we go backwards in time it decreases proportionally to $a(t)$.

This tells us that in the past, at the time when the scale factor was a million times smaller than it is today, the energy

22. One order of magnitude, in physics, means approximately ten, when talking about ratios. So if we add two quantities of orders of magnitude respectively n and $n+1$, with respect to whatever quantity, the result is still of order of magnitude $n+1$.

density in matter and the energy density in radiation were the same.

This is also when the temperature was a million times larger than the 3° it is today.

In other words, when the scale factor was 1000 times smaller than at decoupling, and the temperature was 1000 times hotter than then, there was a crossover point. After that point, matter energy became larger than radiation energy. Before that point, radiation energy was the larger.

And as we go even further and further back, the universe is more and more radiation-dominated.

It tells us that, if we go to the very early universe, massive particles, protons, electrons, some nuclei, were very unimportant in the energy balance. And the Friedmann equation was basically just the equation coming from radiation, see equation (25b) of chapter 2.

Of course that was very early, and we don't easily see directly back to that time. So in fact we don't easily see back to a time when the universe was radiation-dominated. Nevertheless theory tells us that it must have been radiation-dominated.

The date of the crossover point between radiation-dominated and matter-dominated is estimated to be of the order of fifty thousand years after the beginning of the universe.

Before that crossover point, there was an even earlier time when protons did not exist. We will study that.

And of course, all of this is way before galaxies and stars appeared. Indeed they appeared only long after decoupling, which is also the recombination time. We will study that too.

So we got already two landmarks after the beginning of the universe :

a) transition from a radiation-dominated to a matter-dominated universe, *c.* 50 000 years after the beginning of the universe, and

b) decoupling between photons and massive particles, protons and electrons, when those began to be mostly combined into atoms and therefore the universe became transparent, *c.* 400 000 years after the beginning of the universe.

To fix ideas not on a time scale but on an expansion factor scale, at the first landmark, the scale factor $a(t)$ was a million times smaller than it is today. And at the second landmark, it had grown by a factor of one thousand, and was still, in linear dimension, a thousand times smaller than today.

Since, as we said, the universe is estimated nowadays to be 13.8 billion years old, and the decoupling took place roughly four hundred thousand years after its beginning, we see that the period from the beginning until decoupling is tiny. It is negligible on a cosmological timescale. And after that decoupling point, the universe expanded by a linear factor of one thousand. Or in volume, it expanded a billion times.

Some cosmologists assign the name *Big Bang* to the beginning of the universe, others give it to the time when matter began to dominate, others yet when it became transparent. Over the 13.8 billion years of the universe, we see that in terms of time these are not big distinctions. In terms of expansion of course it is different.

If you place the Big Bang at decoupling, the universe expanded by a factor 1000 between the Big Bang and today. But if you place it earlier, the expansion was bigger.

Exercise 2 : If you place the Big Bang at the crossover time when the universe went from radiation-dominated to matter-dominated,

1. **by how much did the scale factor expand between the Big Bang and today ?**
2. **what was the volume expansion of the universe over that period ?**
3. **what was the temperature then ?**

The next important landmark, going backwards in time, is when the temperature was hot enough to create positrons.

Before going into that, let's do a questions / answers session.

Questions / answers session (4)

Question : What was $a(t)$ at time zero? Does it have a meaning?

Answer : Remember that as far as we can see, the universe appears pretty close to being flat. It appears to be essentially a flat 3D universe in expansion. And if it was always flat, it was always infinite, even at the beginning.

In a flat space only ratios of a's have meaning. If you take a flat plane, or a flat 3D space, and you ask : what the radius of curvature of it is? It doesn't mean anything. Or, if you prefer, it is infinite.

But, as we already saw in chapter 1, if the plane or the 3D space stretches by a factor of 2, so that the grid that is embedded in the space stretches by a factor of 2, that is well defined.

So to the extent that the universe is and has always been about flat, $a(t)$ does not really measure the radius of anything. It is ratios of a's that reflect the expansion.

Q. : So at its beginning the universe could have been infinite?

A. : We don't know what it is was like at its beginning. But that is indeed possible.

As said, we believe it is about flat, and has always been so, so that only ratios of a's have a meaning.

To imagine that the universe may have started like a pinhead, I think is an incorrect way to think about it [23].

It really depends whether $k = +1$, 0 or -1 in Friedmann equation (1).

If $k = +1$ that means a closed and bounded universe. The scale factor a then does have some intrinsic meaning. Still, the only thing we know is ratios of a's. We don't know what the primordial size of the universe was, when it first formed, in the case it was a sphere.

If it is negatively curved, then it started out infinite. The scale factor a again is identified to a radius – a hyperbolic radius in this case. But the space is always infinite. We tried to give a feel for what a 2D hyperbolic space is, using a stereographic projection of the Escher drawing, which mapped the entire space onto a disk, see figure 18 of chapter 3. Of course, if $k = -1$, our universe is not a 2D but a 3D hyperbolic space.

Moreover, the universe is in expansion. The factor $a(t)$ increases with time.

23. See chapter 9 for details on that point. If we set the beginning of the universe at the point in time where it was an infinite space, about flat, and empty, except for potential energy – which was converted among other things into high energy photons, uniformly distributed in space, and those in turn produced particles and antiparticles –, then by definition the universe started infinite and flat. Only after this point do we speak of a spatial universe. And from it, the space began its expansion according to the scale factor $a(t)$. But there is a theory, called *inflation*, of what happened before this starting point. Inflation lasted about 10^{-32} seconds. And at the beginning of inflation the universe may have been a tiny 3-sphere.

So, yes, we work from whatever observational cosmologists can say. And what they can say is about ratios of a's.

Q. : So, going backwards in time, when we get to decoupling, is that when our observations end ?

A. : It is where our *optical or electromagnetic observations* end. But that is not where all our observations end. Let's draw a new picture, and discuss that a little bit.

Last lesson, we drew a picture of space vs time, time going upward, space going horizontally, figure 8.

Figure 8 : Representation of spacetime. For convenience space is one-dimensional. I.e. think of events as happening spatially *only on a line*, but they are plotted in spacetime in two dimensions.

In looking back to different distances, we see different time periods. We can plot that by saying we sit at the centre, the point t = today, and $r = 0$. And we will look out.

The new picture we shall draw is of space alone, without a time axis. And since the time dimension is not represented, we can represent two spatial dimensions, figure 9.

We can look at things a thousand light years away in distance. We see things that happened a 1000 years ago in time. It is the smaller shell in figure 9.

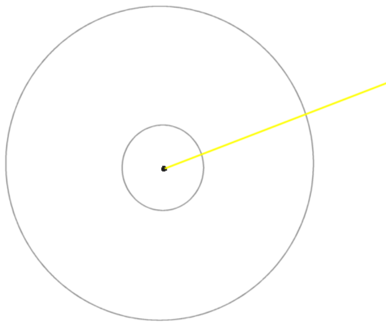

Figure 9 : Two-dimensional space viewed by us sitting at the center.

We can look at things a million light years away in distance. We see things that happened a 1,000,000 years ago in time. It is the larger shell in figure 9. (It is not at scale.) We can look even further out.

Eventually, looking out, we see back to the time at which the recombination took place.

Looking in any direction of the 1-sphere representing the sky around us, we see light rays coming in at time $t = 0$.

As in figure 8, they carry information on events that took place at different times in the past, but arrive at us together at time zero.

So, we see light coming in from different places and that is the subject of astronomy of course.

Before decoupling took place, the universe was ionised. That is the whole point : light scatters a great deal from ionised material. So light didn't make its way very readily through the ionised material before the decoupling. That basically means, roughly speaking, that the universe was optically opaque before decoupling [24].

---

24. It is important to understand that whatever we perceive directly is always *here and now.*

When, at time t = *today* at our clock, we look at some object far away, we actually perceive light emitted by that object at some time in the past, and arriving at us now. We don't *directly* look far away.

We see here and now light coming from there and then. In other words, *we have to change our way to think of how we see the universe.* We don't see instantaneously a large volume. Looking out is somehow looking at rings or shells of events more and more distant not only in length but in time.

Perhaps thinking of only hearing sounds – for example the thunder – can help figuring out the nature of the perception of the space organized into shells corresponding to different times.

Do not mix up what the universe *is* at any time, and *what we see of it* at that time. As said, k appears to be equal to 0. In that case the universe is and has always been flat, therefore infinite. Its scale factor $a(t)$ is increasing with time. But the universe has always been infinite, with things located all over the place, including places too far to be seen by us today.

Hence, when we look at the shell corresponding to the time of decoupling, we are looking at the microwave background basically.

There are telescopes that are looking at the microwave background, effectively looking back to that shell corresponding to decoupling time. Beyond that, they cannot see because it is opaque.

Now we might have a chance of observing other kinds of things coming through, that are less affected by whatever it is that is creating the opacity – for example neutrinos.

Low energy neutrinos can pass through a lot more stuff than can photons. So in principle we can see neutrinos from farther out.

Gravitons, even more so, can go through the stuff that is opaque for photons.

The distance to which we can observe in principle may not be constrained to the same degree as the things that we can see by optical or electromagnetic radiation. With elec-

---

At the beginning, whatever $a(0)$ was – and it has no intrinsic meaning for a flat universe –, the universe was an *infinite flat space* filled with an extremely hot soup of photons, not yet thermalized.

Most of the photons we see today date from the decoupling time. They were in thermal equilibrium and since then are somehow frozen into the blackbody spectrum, with only their temperature decreasing, because it varies like the inverse of $a(t)$. They are all over the place and are called the CMB. We don't see farther than the distance covered by light since decoupling. Yet the universe was larger. It was infinite.

tromagnetic radiation, the earliest that we can see to is the decoupling period. But there are other messengers from further back.

Anything beyond decoupling, however, is only circumstantial evidence. Some of it we understand very well though.

Let's close the questions / answers session, and keep going back in time.

Primordial soup of electrons, positrons and photons

We explained that, in figure 9, there is a largest shell corresponding to things that we can see. We can see today the photons they emitted at some time in the past. We can see them optically or with electromagnetic radiation. And that largest shell corresponds the decoupling time.

But there are things further back that we can occasionally observe too, not from the electromagnetic radiation, i.e. the photons they emitted, but from other telltales like neutrinos or gravitons. Times earlier than decoupling correspond to shells larger than the decoupling shell.

For example, we can think of a shell corresponding to the transition period from radiation-dominated to matter-dominated, figure 10.

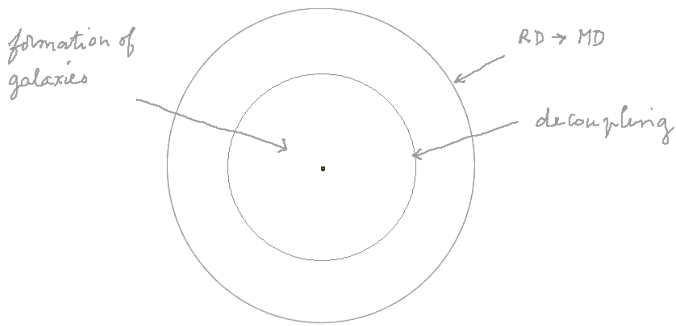

Figure 10 : Shells corresponding to various landmarks. (Scales are not respected.)

Sometime between decoupling and now, galaxies formed. We will study their formation.

Thus figure 10 is a picture of the universe in a nutshell.

More things can be traced back yet. We understand the laws of physics well enough to go to even higher temperatures. It is an issue to go to ever higher temperatures.

In figure 10, despite the fact that it looks like things are getting bigger as we go out, the scale factor itself is getting smaller. Moreover, on the scale of 13.8 billion years which is the estimated age of the universe, the outer shells we are looking at, which correspond to a few ten or hundred thousand years after the beginning, should be close to each other. So linear scales are not respected in any way.

When the scale factor is getting smaller, according to equation (21) temperature is increasing.

Something new happens when the temperature reaches ten billion degrees Kelvin. That corresponds to

$$\frac{a_{today}}{a(t)} = \frac{T(t)}{T_{today}} = \frac{10^{10}}{3} \approx 3 \text{ x } 10^9 \tag{33}$$

In other words, the temperature was 10^{10} Kelvin when the scale factor was three billion times smaller than it is today.

The characteristic photon energy at that temperature can be calculated. From formula (15), which gives the wavelength of the characteristic photons, that is those corresponding to the peak of the blackbody spectrum,

$$\lambda = \frac{hc}{k_B T}$$

and $\lambda\nu = c$, we have that

$$\nu = \frac{k_B T}{h}$$

where $k_B = 1.381 \text{ x } 10^{-23}$ J/K, $h = 6.63 \text{ x } 10^{-34}$ Js, and the units of ν are 1/s, i.e. the inverse of time.

This yields for the frequency of the characteristic photons at ten billion degrees Kelvin

$$\nu \;\approx\; 2 \text{ x } 10^{20} \; s^{-1} \tag{34}$$

Then, according to the formula $\epsilon = h\nu$, and using the conversion 1 $eV = 1.6 \text{ x } 10^{-19}$ Joule, we get for the energy of one photon

$$\epsilon \approx 1\ MeV \tag{35}$$

This is approximately twice the energy of an electron [25].

So, once we get to high enough temperatures, there are lots of photons around whose energy is such that if they collide they can make a transition to an electron and a positron [26] – also called a positively charged electron.

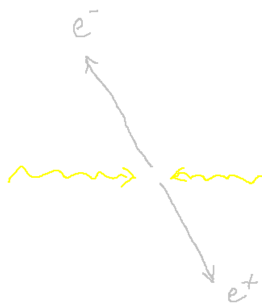

Figure 11 : Two photons colliding and producing an electron and a positron.

Whether such a collision does or doesn't happen is a matter of quantum electrodynamics and computation. But it could not happen when the energy of the characteristic photons is much lower than the energy mc^2 of the mass of an electron. It could still happen for two photons in the tail of the

25. Remember that the energy of a proton is 1 GeV, and an electron has 1/2000 the mass of a proton. Therefore, two electrons have 1/1000 the mass of a proton, and an energy of 1 MeV.

26. A temperature somewhat lower than ten billion degrees Kelvin would be sufficient. We just need each photon in the collision to have the energy of one electron. Therefore two photons can produce two electrons, one ordinary negatively charged and one positively charged.

spectrum, but remember that the spectrum falls very fast. Most of the energy of the soup of photons is around the caracteristic frequency, or wavelength, which is itself simply linked to temperature.

So once we get up above the threshold of 0.5 MeV per photon [27], they have enough energy so that when they collide they can make electron-positron pairs. And they do.

QED [28] says the photons of such energy, when they collide, have a significant probability of creating an electron-positron pair.

It is also true that electron-positron pairs will collide together and make pairs of photons. What happens is the electrons, positrons and photons come into thermal equilibrium.

The number of electrons and positrons is no longer determined by memory of the future – memory of the future meaning today. We come to a point where new particles can legitimately be created by simply the energy of collision of photons. And, in the equilibrium, the number of such electrons and positrons is determined not by history, but just by the temperature.

In fact, when you are in thermal equilibrium at that temperature of ten billion degrees, the number of electrons is equal to the number of positrons – not quite, we will see in

27. We no longer remind the reader that we are talking about *caracteristic* photons, i.e. those at the peak of the spectrum, because most of the energy is carried by photons around that frequency.

28. QED is the abbreviation of quantum electrodynamics.

a moment, but close.

The number of electrons is about equal to the number of positrons, and is also about equal to the number of photons.

$$N_e \;\approx\; N_{e^+} \;\approx\; N_\gamma \tag{36}$$

In other words it is a soup consisting of the three types of particles in equal numerical proportions.

What happens if the number of electrons and positrons gets too low ? It is replenished by collisions of photons.

What happens if the number of electrons and positrons gets too high ? They will collide and make photons.

So there is a balance. And it is not determined by anything except the temperature. At 10 billion degrees, the three species of electrons, positrons and photons are all equally abundant.

Remember that the number of photons N_γ today is about 10^8 times larger than the number of electrons today. The implication is this : at the time of the primordial soup of electrons, positrons and photons, the following ratio held

$$\frac{N_e - N_{e^+}}{N_e + N_{e^+}} \approx 10^{-8} \tag{37}$$

Let's explain why. The numerator on the left-hand side is the *excess* of electrons to positrons. That number doesn't change. It was the same at the time of the primordial soup, as it is today.

In fact, the number of electrons minus the number of positrons is basically today's number of electrons, whatever that number is.

As the universe cools off, what does change is the *sum* of the numbers of electrons and positrons, that is the denominator on the left-hand side of equation (37).

The number of electrons plus the number of positrons, at the time of the soup, was twice the number of photons N_γ then.

So the ratio on the left-hand side of equation (37), at the time when the universe was hot enough to create lots of electrons and positrons, is believed to have been the same [29] as N_e/N_γ today, that is 10^{-8}.

If we forget about today, and simply ask : what was life like at the very early times when electrons and positrons were as abundant as photons ? The answer is : basically for every 10^8 electron-positron pairs, there was 1 excess electron.

Now that is an extremely odd fact. It turns upside down the old question of why there are so many photons. The question is not why there are so many photons today, as it is why there are so few electrons.

At the time of the primordial soup, why was the excess of electrons over positrons so much smaller than the total number of electrons and positrons ?

29. To be precise $N_e + N_{e+}$ of then, is equal to twice N_γ of today. But that doesn't change the order of magnitude of 10^{-8} for the ratio (37).

In other words, the question is not how the number of photons gets to be large. It is why the excess of minus charges over plus charges is so terribly tiny.

But that question can even be turned around again by asking what is it that lead to *any* difference between the number of electrons and the number of positrons.

If the universe really started neutral – and for the moment suppose there were no protons, just lots of photons, lots of electrons and lots of positrons –, and if we believe in symmetry, in particular symmetry between particles and antiparticles, then we might guess that the way the world started was with an equal number of electrons and positrons.

If it had no electrons and positrons to begin with, that would change quickly. The photons would collide with each other, and make electrons and positrons in equal numbers.

So a natural expectation, based on symmetry, would be that the number of electrons and the number of positrons was equal.

Looking at the ratio

$$\frac{N_e - N_{e^+}}{N_e + N_{e^+}}$$

the denominator would be some big number, but the numerator would be exactly 0.

That turns the question around again. Given that the number of electrons and positrons are not equal – which is odd in itself – the question then becomes : why is the difference so small ?

But, on the other hand, you could also say : well, 10^{-8} is a lot bigger than 0, so why is it so big ?

Why is it so big ? Why is it so small ? What is a theory which tells us why there is an electron-positron imbalance ? We will address that question in the next chapter. It is the theory of the excess of particles over antiparticles.

For now, let's push further back in time, another factor of 100, or 200, or 300 in terms of temperature, or inversely in terms of scale factor. Something else then begins to happen.

Quarks and antiquarks

If we go back by a factor of 1000 – probably a good deal less than that – then the photon collisions have enough energy to make protons and antiprotons.

In other words, when it is a thousand times hotter than in the previous section, the universe should be filled with protons and antiprotons, again in equal abundance.

The trouble is it is too hot for protons and antiprotons to exist. It is simply the same phenomenon as when the hydrogen dissociates into electrons and protons. The protons and neutrons – the protons in particular – would dissociate into quarks.

So above 10^{12} K or 10^{13} K, the world is a hot soup of electrons, positrons, photons, quarks and antiquarks.

And let's not forget dark matter, which we don't know exactly what it is, and gluons, and all sorts of things. The population gets more and more complicated as we try to be more complete.

Shouldn't we also add dark energy ? No, dark energy is not important in the early universe. It is not part of the story at this stage.

But what is true is there are some relics. We are dinosaur hunting. What are the things which didn't change since the very earliest times ?

It is expected, or thought, that the excess of matter to antimatter, in particular electrons to positrons, goes way back.

Similarly, we think that there was an excess of quarks to antiquarks, because there are more quarks today than there are antiquarks. We know that because there are protons and not many antiprotons out there today.

So in the hot soup of electrons, positrons, photons, quarks and antiquarks we are concerned with, there is an excess of quarks over antiquarks. And there is an excess of electrons over positrons. The question is : what is the origin of that excess, and why is it so small ? What is the theory that accounts for it ?

We speak of a single excess, because the two excesses are linked. They are the two sides of the same phenomenon.

The excess of baryon matter to antibaryon, which means quarks to antiquarks, is the same as the excess of electrons to positrons [30]. That is to keep electrical neutrality.

We can ask : what in our experience with the physics could account for that imbalance ?

We do have a theory of it, not a detailed quantitative one though. The number 10^{-8}, in equation (37), has never been calculated as a hard number. But there is a rough theory of it. We more or less understand why it is small.

Questions / answers session (5)

Q. : Did the quark combination into protons and the electron-proton combination into hydrogen atoms happen at the same time ?

A. : We use the temperature or the scale factor to classify events, not the time chronology. Quark combination happened way back before electron-proton combination. The universe was much hotter than the 4000° K of the recombination or decoupling.

Let's go over recombination again. We start with this condition : there is lots of electrons and positrons, lots of photons. They make a hot soup. It is almost completely balanced between electrons and positrons, with a tiny excess of the

30. Since a proton is made of three quarks (2 up quarks and 1 down quark), there is an extra factor 3 in the formula for the quark-antiquark imbalance matching the electron-positron imbalance.

former. And let's follow it forward in time.

As we follow it forward in time and it starts to cool down, the electrons and positrons will annihilate into photons.

But typically the photons will not have enough energy to collide and make electrons and positrons again. So as you cool down, annihilation of electrons and positrons becomes the dominant thing that happens. The electrons and positrons just annihilate. They just disappear and turn into photons.

The photons scatter. And when they scatter, they thermalize. And the temperature comes below the temperature needed to make electrons and positrons.

If the number of electrons and positrons were equal, what would simply happen is essentially all the electrons and positrons would annihilate and there would be none left or almost none left.

But what we see is the leftover because it wasn't exactly balanced. We see the imbalance.

And, as far as we can tell, electrons dominate over positrons *everywhere in the universe*. It is not as if we were just a local pocket where there is more electrons than positrons – as if it were a statistical fluctuation.

No matter how far out we look, what we see appears to be due to electrons and not to positrons. We see cosmic rays coming in. The highest energy ones are really cosmic in origin. They come from very far away. Yet nobody has ever seen an antinucleus come in in cosmic rays.

We observe a few antiprotons, but that is easy to explain. When high energy particles come in, they hit the atmosphere, and it will make some antiprotons. But the chances of them making an antinucleus – a whole helium antinucleus – are extremely small. So if we observed some, they would be coming from very far away.

On the other hand, if there were galaxies out there that were antigalaxies, then we should see coming from them antinuclei, that is nuclei of the opposite charge than the nuclei we see coming from ordinary matter galaxies.

So the complete absence of antinuclei in cosmic rays is pretty strong evidence that everywhere throughout the universe $N_e - N_{e^+}$, the excess of electrons over positrons, has the same sign, and is of the same order of magnitude. And in fact it is of the same order of magnitude.

So there is something to explain then.

Q. : $N_e - N_{e^+}$ is also the number of protons ?

A. : Yes. At the time of recombination the protons are still the same protons. Proton pair creation has not happened yet [31].

31. Caveat : here we look at the chronology *backwards*. When we say "something has not happened yet", we mean that "it happened earlier in time". We have to go farther in the past to meet it.

So at the time of recombination $N_e - N_{e^+} = N_p$.

Q. : What was the proton population at that time?

A. : At that time the protons were pretty much just the same protons that exist today.

They were kicking around. The temperature was high enough to keep them kicking around a lot.

But it wasn't high enough to destroy them. And it wasn't high enough to create new proton-antiproton pairs.

So there was a period when the protons were just plain old protons. And there was a tiny population of them compared to the photons, a ratio of 10^{-8}. And the proton number was the same as the difference between electrons and positrons.

Q. : Could not the imbalance between electrons and positrons be just due to randomness in the limited region of the universe that we can see?

A. : The imbalance is much too significant to be due to a local statistical fluctuation.

Local statistical fluctuations over of volume of, let's say, 10^{10} light years cube, all of which have the same sign of imbalance, and the same order of magnitude, such random statistical fluctuations of matter over anti-matter would be violently unlikely.

So the fact that over the 10 billion light years or the 20 billion light years or so that we see it is all of the same sign certainly needs an explanation. And the explanation cannot be a random statistical fluctuation.

But the point is correct. If we simply can't see out beyond a certain distance, and there is some kind of reason why there might be a transition to another region where everything is backwards with respect to electrons and positrons, we wouldn't know it easily.

So the answer is : we don't have much information about what is out there beyond 10 or 20 billion light years.

Q. : Could there be galaxies beyond the farthest distance that we can see ?

A. : Yes there could, and for the following reasons there probably are.

Beyond the farthest distance that we can see we don't really know what there is – because we cannot see what there is.

We see up to the CMB formation, which took place at the time of decoupling. Beyond that the universe is opaque to us. The galaxies that we can see formed after decoupling, figure 10.

Whether there are galaxies up beyond the decoupling distance, we cannot tell. We could not see them because we cannot get light from them. From too great distances we just don't get light altogether.

But certainly we can get light from comparable times, even though figure 10 tells us nothing about whether there are galaxies out beyond the CMB.

What we do know is that by looking out as far as we can, we see galaxies homogeneously distributed. There is no reason to think that there is a sharp edge of any kind in the galaxy repartition in the universe. It just looks uniform.

So most people would bet that this uniformity extends out beyond the distance that we can see through.