Lesson 6 : Dark matter and allocation of energy density

Notes from Prof. Susskind video lectures publicly available on YouTube

Introduction

In this lesson we come to some observational facts. More precisely we come to connections between the observational facts and the equations. It is impossible to give any kind of justice to observations without putting them into the context of the fundamental cosmological equations [1]. So, we have already talked about them, but we begin this lesson with observational facts.

Let's define some observational quantities – quantities which astronomers can and do observe. The basic equation in which they appear, as always, at least for a homogeneous universe, is the Friedmann equation. Here it is in one of its simpler forms

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{1}$$

The left-hand side is the square of the Hubble constant $H = \dot{a}/a$. Remember that it is constant over space, but it may vary over time. Usually when we talk about the Hubble constant without further specifications, we talk about its value today.

On the right-hand side of equation (1), we have the good old factor $8\pi G/3$ in front of the usual energy density ρ, minus a term coming from curvature.

1. Indeed, in science in general, observations make sense only within the framework of a theory to interpret them. Phenomenology tells us that there is no such thing as pure observations – with the possible exception of extremely raw perceptions, and even that is debated. These are the ideas underlying the statement that we don't *observe* the world, we *construct* it.

The curvature term has a k in front of it. It can be either $+1$ or -1 or 0, and nothing else.

a) 0 for flat geometry,

b) +1 for spherical geometry,

c) −1 for hyperbolic geometry.

To work with Friedmann equation, a preliminary step is to express $\rho(t)$ in terms of $a(t)$. We did it in various ways. In particular, to that effect, we introduced the thermodynamic *equation of state*, see chapters 4 and 5. Then we have to solve Friedmann equation which is a differential equation where the unknown function is $a(t)$, which for short we often just denote a.

The role of a is to go into the metric. The metric is

$$ds^2 = -dt^2 + a(t)^2 \ [\ dr^2 + \text{an angular piece}\ ] \tag{2}$$

The angular piece is one of three possibilities depending on k, that is on the geometry of the universe. The first case is flat space, $k = 0$. Then in equation (2) the term in brackets is

$$dr^2 + r^2 \ d\Omega_2^2 \tag{3a}$$

where $d\Omega_2^2$ is the metric of the 2-sphere of the sky.

The second case is spherical geometry, that is $k = +1$. Then the term in brackets is

$$dr^2 + \sin^2 r \ d\Omega_2^2 \tag{3b}$$

The third case is hyperbolic, $k = -1$. In that last case the term in brackets is

$$dr^2 + \sinh^2 r \; d\Omega_2^2 \tag{3c}$$

Instead of writing three separate equations, we just call $\xi^2(r)$ the function in front of $d\Omega_2^2$. Then in the three cases, the metric has the same form

$$ds^2 = -dt^2 + a(t)^2 \; [\; dr^2 + \xi^2(r) \; d\Omega_2^2 \;] \tag{4}$$

where

a) $\xi = r$ if $k = 0$,

b) $\xi = \sin r$ if $k = +1$,

c) $\xi = \sinh r$ if $k = -1$.

That makes one simple formula. We just have to remember which one we are using. So we have three different spatial geometries, and one equation of motion for $a(t)$, equation (4). That is our basic framework.

Now the energy density ρ, in equation (1) and its variations, can come from various sources. The actual only sources that we really know in the current universe are of three kinds. In other words there are three contributions to ρ.

These three kinds of contributions replace the term $8\pi G/3$ times ρ in equation (1). It can be rewritten as an addition

$$\frac{8\pi G}{3}\rho = \frac{c_R}{a^4} + \frac{c_M}{a^3} + \Lambda \tag{5}$$

In the past lessons, we studied in turn each of the three terms on the right-hand side.

The first is the *radiation* energy density. It is some constant c_R, representing the number of photons per unit volume, or the number of units of energy per unit volume, when a was equal to 1, divided by a^4.

Then there is the contribution associated with ordinary non-relativistic particles. We usually call that *matter*. It is a coefficient c_M divided by a^3.

And there is the possibility of *vacuum energy* density. It is denoted Λ, the cosmological constant. As we saw it doesn't dilute with increasing radius or scale factor a.

We can add in the curvature term $-k/a^2$, as part of ρ, on both sides of equation (5), to get the complete right-hand side of Friedmann equation.

$$\frac{8\pi G}{3}\rho - \frac{k}{a^2} = \frac{c_R}{a^4} + \frac{c_M}{a^3} + \Lambda - \frac{k}{a^2} \tag{6}$$

The curvature term can be thought of as a kind of energy if we like : the energy from curvature, scaling like $1/a^2$. And all of this is equal to the square of the Hubble constant.

$$\frac{c_R}{a^4} + \frac{c_M}{a^3} + \Lambda - \frac{k}{a^2} = H^2 \tag{7}$$

Now let's think about this equation today, at the present time. Indeed, a depends on time, and therefore so does H. It says that a certain set of four quantities has to add up to H^2. This can be viewed as a constraint.

Now we can ask : in the present universe, on the left-hand side of equation (7), how much comes from radiation ? How much comes from matter ? How much comes from dark energy, i.e. from lambda ? And how much is left over in the curvature term ? That is a reasonable question.

The value of the scale factor today is some number which we are going to talk about in a moment. We will see how big it is ?

So, today, each of the four terms on the left-hand side of equation (7) has some value. And H has some value. Let's think about how we measure them.

H is the Hubble constant. Hubble [2] measured the Hubble constant in 1929. He did that by measuring the relation between velocities and distances – velocities determined from redshift, and distances from luminosity. A distant bulb looks dimmer than a close bulb. So he basically plotted redshift versus luminosity and found a more or less linear relation because the measurements essentially concerned a fixed time t, namely today [3].

The fact that he mismeasured H by a factor of 10 is not the important point. In principle, with better measurements, he would have got a good estimation of H today. These were not cosmological measurements, incidentally. By cosmological we mean measurements taking place on a scale of tens of billions of light-years. They were measurements of the lo-

2. Edwin Hubble (1889 - 1953), American astronomer.

3. The Belgian Jesuit and physicist Georges Lemaître (1894 - 1966) had predicted such a linear relationship for a fixed time, on theoretical grounds from Einstein general relativity, in 1927.

cal properties of a few clusters of galaxies. So they did not penetrate back deep into the past. For his purposes, since the measurements weren't penetrating back deep into the past, he wasn't looking at his constant at different times, he was looking at it now.

When we say that distances are measured by luminosity, we gloss over a complicated search. There is a whole story of the cosmic distance ladder. It is a story of many different little pieces overlapping. How do you measure the distance to the nearest stars ? You do it by parallax. Then you find that certain stars always behave in a particular kind of way. These stars are called the *Cepheid variables*, first observed in the Cepheus constellation, between Cassiopeia and Ursa Minor.

The Cepheid variables display a strong relationship between their luminosity and their pulsation period. These measurements are determined from stars which are close enough, that we can get a good evaluation of their distance from parallax. Of course they still are in the Milky Way. Then for a given pulsation, luminosity becomes a more accurate measure of distance.

In other words, the Cepheid variables provided us with a standard candle to work with. We could then employ it to measure the distance of things farther away. Thus we worked our way up a ladder. Supernovae are the best standard candles. We are going to talk about that.

So Hubble measured velocity by redshift, and distance by a sequence of overlapping standards that were available at his time. Today we do a better job of measuring H. But basically we measure H by measuring the relationship between velocity and distance, or redshift and luminosity. When we want to do it, we don't need to go to very huge distances to do so. Therefore we are not looking back deep into the past, and we don't have to worry too much about H varying with time.

Now what about c_R (read *c radiation*) ? We can actually look at the amount of radiation in the universe. Most of it is in the form of the Cosmic Microwave Background, which we will come to. But it is very feeble. There isn't much energy out there in the form of photons. It is a tiny fraction of the other kinds of energies. We know that simply by measuring the photons that exist around us, the black body photons, the ordinary photons. And again there is nothing particularly cosmological about the measurement. We don't have to go very deep into the distant past nor do we have to go very far away. We have to have a big enough volume to average over though. But we pretty much know that *today* radiation is inessential, is not very important.

Let's come to c_M (read *c matter*). What do we know about matter ? We look at the galaxies. We see how much hydrogen is in them. We see how much luminosity is coming out of them. We make studies of stars. We know a good deal about average stars. And from the average properties of stars and interstellar gas, and things which are more technical but not conceptually difficult, we measure all the mid matter that we can see – the luminous matter [4]. And we add it all up.

4. By luminous matter we mean matter that *radiates*, not necessarily in the visible range of electromagnetic waves, but which does emit radiation.

As the years go on, we measure better and better how much luminous matter is out there. We have a pretty good idea of it now.

The luminous matter is presumably all in the form of atoms and protons and nuclei and electrons, etc. And we get an estimate of the number c_M in equation (7). That number divided by a^3, the density of energy in the form of matter in the universe, today on the average is about one proton per cubic meter. So it is very low, but much higher than the radiation energy density.

We have taken care of the first two terms on the left-hand side of equation (7). And we have taken care of its right-hand side. We will see how we measure a in a moment. Let's continue.

For the moment, let's forget about the vacuum energy density Λ. We will also come back to it.

I want to be slightly historical in the presentation of how we relate Friedmann equation to observations. So I'm explaining cosmology as it was when I was a student in the 1960's : c_R/a^4 was negligible. c_M/a^3 was substantial, one proton per cubic meter. Nobody believed there was a lambda. So we omitted lambda. And H had been measured.

By the way, we often work with the percentages that each kind of energy respectively contributes to the total H^2. These percentages are denoted with the letter Ω with different subscripts. We have

$$
\begin{aligned}
\Omega_R &= \frac{c_R}{a^4 H^2} \\
\Omega_M &= \frac{c_M}{a^3 H^2} \\
\Omega_\Lambda &= \frac{\Lambda}{H^2} \\
\Omega_K &= \frac{-k}{a^2 H^2}
\end{aligned}
\tag{8}
$$

The last two may be positive or negative. And of course they satisfy

$$
\Omega_R + \Omega_M + \Omega_\Lambda + \Omega_K = 1 \tag{9}
$$

As I said, when I was a young physicist, everybody thought that Ω_R was equal to essentially zero, which it is. Ω_M was what we measure when we look at the light coming from the sky, from stars, interstellar matter and so forth, roughly one proton per cubic meter today, divided by H^2. Ω_Λ nobody had ever heard of it. We didn't believe it existed. Einstein had told us it didn't exist. So we then had a situation where we could compute and know what Ω_K was, since the sum of the omega's had to add up to 1.

Or equivalently, going back to equation (7), the term c_R/a^4 was negligible. c_M/a^3 was believed not to be negligible. Λ was unknown and not there. And H had been measured.

The value of H^2 was estimated to be about twenty times or thirty times the matter term c_M/a^3. In other words, H^2 could not possibly correspond to the matter contribution in equation (7). The only conclusion was that it had to come from Ω_K, and that Ω_K was almost equal to 1. Minus k over a^2 had to be close to H^2. First of all that immediately told us that we were in the case $k = -1$. The universe was an open hyperbolic universe. Furthermore the following interesting relation held

$$\frac{1}{a^2} \approx H^2 \tag{10}$$

which in turn gave us the value of a in terms of H. That is what we thought in the 1960's.

Now H has another meaning at least within the context of matter-dominated universe. Let's go back to the matter-dominated universe, which after all is what we have in equation (7) when we eliminate c_R/a^4 and Λ. Here it is

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{c_M}{a^3} - \frac{k}{a^2} \tag{11}$$

Remember, in the case $k = 0$, the solution of this differential equation for a. We saw in chapter 1 that in the flat matter-dominated universe

$$a \sim t^{2/3} \tag{12}$$

We stressed that Newton could have built and investigated this model. It was not dependent on having developed first general relativity, even though people figured it out only after and from Einstein's work. Newton could have found formula (12) for the evolution of scale parameter a as a function of time.

When the universe was young and a small, k/a^2 was smaller than c_M/a^3. At late times it got bigger. But at early times, for most of the history of the universe, even though the matter was pretty small, on the right-hand side of equation (11) the matter term competed favorably with the curvature term. So formula (12) was a pretty good approximation. It fails at late times, but it is pretty good at early time[5].

So let's calculate $\dot{a}/a$. From formula (12) we get

$$\dot{a} = \frac{2}{3}\ t^{-1/3}$$

Then, dividing by a, this becomes

$$\frac{\dot{a}}{a} = \frac{2}{3}\ t^{-1/3}\ t^{-2/3}$$

or more simply

$$\frac{\dot{a}}{a} = \frac{2}{3t} \tag{13}$$

So, in this model of a matter-dominated universe, we know how the Hubble constant varies with time : it is inversely proportional to time. Of course, as we have often repeated it is not a constant

$$H = \frac{2}{3t} \tag{14}$$

5. As usual, we use the reasoning that two differential equations, close to each other in some sense, must have solutions close to each other. It can easily be justified mathematically in the cases where we use it.

It is not a constant, but we can ask what is it today. Let's call it H_{today}. Then H_{today} is two thirds of the time today. What is the time today ? It is the age of the universes. Let's denote it T.

$$H_{today} = \frac{2}{3T} \tag{15}$$

In other words, how long it was since the universe was very small ? What exactly we mean by very small is not important. How long it was since the scale factor was many orders of magnitude smaller than today ? That is what H is. It is one over the age of the universe, or to be more exact two-thirds over the age of the universe.

The units of the age of the universe, T, are the units of time, seconds, years, whatever we want to use. Notice that there is no speed of light $c = 1$ hidden anywhere in formula (15). And the units of the Hubble constant are one over time. If time is measured in years, the Hubble constant is approximately $\frac{1}{2}$ x 10^{-10}.

The Hubble constant is $\dot{a}/a$, so astronomers tend to express it in awe-inspiring units of some astronomical speed divided by some astronomical distance, because it is meaningful for them when using telescopes. But from our theoretical point of view the Hubble constant, in the matter-dominated model, is just one over the age of the universe, multiplied by 2/3. And it is mesured in the inverse of time units.

Incidentally, what would it be in a radiation-dominated model ? We saw that formula (12) would become

$$a \sim t^{1/2} \tag{16}$$

And we would get

$$H_{today} = \frac{1}{2T} \tag{17}$$

In fact, for most equations of state, the Hubble constant today is related to the inverse age of the universe.

So we learned more about the Hubble constant, having measured it. Hubble measured it. As we said, the value he got for H was too big by a factor ten. Consequently his estimate of the age of the universe was too young. He thought it to be roughly one billion years old, while our estimate today is more like 10 billion years [6]. But that is just from measuring Hubble constant.

Let's go back to equation (10) and consider it to be an exact relationship. Furthermore let's assume for simplicity that H today is just $1/T$. So we have

$$\frac{1}{a^2} = H^2 = \frac{1}{T^2} \tag{18}$$

In other words, in the simple model derived from equations (1) to (10), the size of the universe is basically the age of the universe. Now there is the speed of light in equation (18). a is measured in length, T is measured in time. Therefore in the equation

$$a^2 = T^2 \tag{19}$$

there is a factor c^2 in front of T^2 to make the units jibe. And in our calculations the units are such that $c = 1$. It is

6. The best estimate, in 2017, is $13,8$ x 10^9 years.

left as an exercise for the reader to keep track of the speed of light c in Friedmann equation (1) and in the subsequent calculations.

Thus cosmology 40 years ago said that the size of the universe is about equal to the distance light will travel in the age of the universe. It is what was expected to be true on the basis of the knowledge available then. Let's stress that our vision has since changed.

But the model we described explains why, for all these years, people thought that the radius of curvature of the universe was about equal to its age – something that we no longer think is true. We now introduce also dark matter and dark energy into Friedmann equation.

So the first change has to do with dark matter.

Dark matter

To explain what dark matter is, let's begin with ordinary matter.

Ordinary matter, made out of electrically charged particles, radiates. It radiates when it is accelerated. It radiates when you collide with it. And it is because it radiates that, be it in the visible range or the invisible one, it is called luminous.

It is a kind of matter that we know about from direct telescopic or radiotelescopic observation. And it is represented by the term c_M/a^3 in Friedmann equation [7].

On the other hand, there is another kind of definition of matter : it is something that creates a gravitational field, and is affected by a gravitational field [8]. In short we say it is something that gravitates.

What is the mass of matter which gravitates in the universe ?

At first it was thought that the galaxies were made of luminous matter. The mass of a galaxy was just the mass of its luminous matter.

In 1932, however, Jan Oort [9] noticed that something was wrong with the way galaxies behaved. A year later Fritz Zwicky [10] noticed that clusters of galaxies were misbehaving too. By misbehaving we mean that their gravitational behavior appeared inconsistent with Newton's laws, as well as with Einstein's laws.

Oort conjectured that there was more matter out there than had been accounted for in c_M so far. We shall recount the story.

---

7. It is thanks to photons that we see ordinary matter. Nevertheless we are talking about c_M here, not c_R.

8. Of course light itself is affected by gravitation. It is a consequence of the equivalence principle, see volume 4 of the collection *The Theoretical Minimum* on general relativity. But free photons don't create gravitational fields themselves.

9. Jan Hendrik Oort (1900 - 1992), Dutch astronomer.

10. Fritz Zwicky (1898 - 1974), Bulgarian born Swiss astronomer who worked most of his life at Caltech in Pasadena, California.

Question : What is exactly accounted for in luminous matter ? What about black holes ? Answer : In c_M are included galaxies, instellar gases, black holes, etc. Of course it is a bit embarrassing to count black holes into luminous matter. But it doesn't matter whether we count them as luminous matter or not because they are a small fraction of the whole luminous matter. The black hole at the center of the Milky Way is somewhere between one million and one billion solar masses, compared to the total mass of the galaxy which is half a trillion solar masses.

Let's stress right away that black holes are *not* a part of dark matter. For instance the big black hole of the Milky Way is at the center of our galaxy, whereas we shall see that dark matter is not concentrated at the center of galaxies. Oort was the first to show that, for each galaxy, it is more like a very large halo around it.

The argument leading to hypothesize the existence of dark matter has to do with the so-called rotation curves of galaxies.

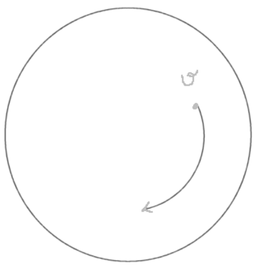

Figure 1 : Trajectory and tangential velocity v of a star at radius r in the outer region of a galaxy.

We look at stars at different radii in the outer region of a galaxy, figure 1. And we look at their tangential velocities. We begin by assuming something which would follow if all of the matter was luminous. The fact is that most of the luminous matter in the galaxy is near the center of the galaxy. So, for all gravitational purposes, the stars in the outer region are moving under the influence of a central force field due to an approximately fixed central mass M. This turns out to be wrong, but let's follow the logic anyway.

Let's apply Newton's equation $F = ma$ to a star at radius r. The magnitude of the centripetal force F attracting the star is

$$F = \frac{mMG}{r^2} \tag{20}$$

where M is the mass at the center of the galaxy, m is the mass of the star, and G is Newton's constant.

From high school physics we remember that the magnitude of the acceleration of a body following a circular orbit of radius r with constant speed v is

$$a = \frac{v^2}{r} \tag{21}$$

So Newton's law writes

$$\frac{mMG}{r^2} = \frac{mv^2}{r} \tag{22}$$

or equivalently

$$\frac{MG}{r} = v^2 \tag{23}$$

This would then say that the velocity falls off like $1/\sqrt{r}$.

It is actually one of Kepler's planetary laws. It is usually expressed in terms of the period rather than the velocity, but it is the same. In particular, for planets on circular orbits, Kepler's third law tells us how the velocity of the planet varies as a function of the radius from the Sun. It varies like the inverse of $\sqrt{r}$.

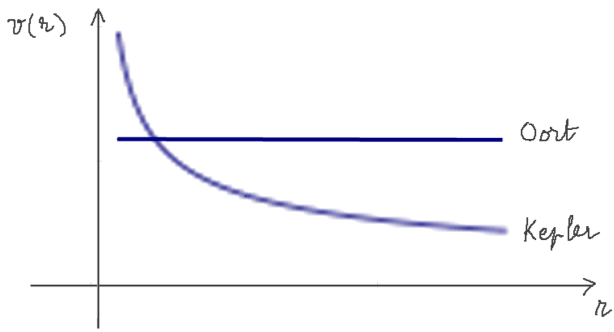

Figure 2 : Tangential speed of stars as a function of r.

However that is not what Oort, nor anybody since, observed for the stars on the outer part of our galaxy. What Oort saw is that the tangential speed is pretty much independent of r, figure 2.

It should be noted that we are not talking about the angular velocity. It is not as if the galaxy were rotating like a giant pinwheel[11]. We are talking about the linear velocity. And this linear velocity, along the tangent to the circular orbit of

11. Neither in Kepler's laws nor in Oort's observations is the angular velocity independent of r.

the star in figure 1, appears to be constant irrespective of r.

So how do we account for that ?

Well, Newton would probably have accounted for it as follows. He would have said : look, there is just more mass out there. So instead of saying all the mass of the galaxy is at the center, let's say its mass is distributed in space. And let's introduce a function $M(r)$.

$M(r)$ stands for the amount of mass contained within a sphere of radius r. Now a star at distance r moves under the influence not of a constant mass M at the center, as expressed in equation (22), but a varying mass $M(r)$. Equation (23) becomes

$$\frac{M(r)\ G}{r} = v^2 \tag{24}$$

If v is constant, we deduce that $M(r)$ must grow like r.

$$M(r) \sim r \tag{25}$$

That doesn't mean that the mass density is constant, incidentally. The local mass density decreases like $1/r^2$. It is left to the reader to prove it.

Exercise 1 : Prove that if $M(r) \sim r$, then the local mass density decreases like $1/r^2$.

And so does the average mass density in the sphere of radius r.

$M(r)$ is the total mass within a radius r. And it apparently increases linearly with r out to the outer boundaries of the galaxy. That is what the application of Newton's law implies.

In fact, it appears that $M(r)$ increases like r even way beyond the visible outer boundaries of the galaxy. Indeed we see things at the far edges of galaxies, stray stars here and there, and they still conform to Oort observations.

Furthermore, even from the motion of stars perpendicular to the plane of the galaxies, it looks as if the distribution of this mysterious mass varying with r is spherical in character, rather than conforming to the odd spiral or flattened shape of galaxies.

For sure, why galaxies look like flattened discs, or have spiral arms, is quite puzzling. The various tentative explanations one can find in the literature are all very complicated. It is not simple, and it is certainly not that galaxies are going around like pinwheels. They are not doing that at all.

So what do we know about this dark matter ?

Just from the fact that it seems to be more or less spherically distributed, and that $M(r)$ grows as r out to distances that are almost midway between neighboring galaxies, one can estimate – by now rather accurately – how much dark matter there is.

This dark matter is of a non luminous variety. And, in the universe, there is about five times more dark matter than ordinary luminous matter. And it is quite bigger in size, see figure 3.

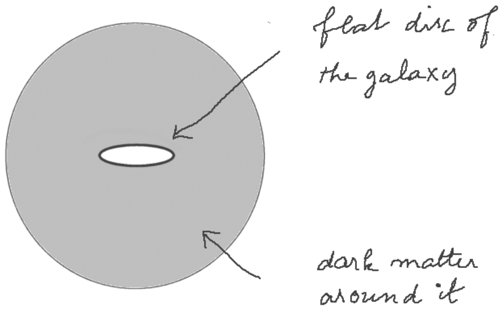

Figure 3 : Galaxy and its spherical halo of dark matter.

Why doesn't the dark matter collapse ? In other words, why didn't it also fall together, during the formation of the universe, the same way that the luminous mass did ?

The galaxies and the stars are the result of the collapsing of ordinary matter. We will come back to this phenomenon. Ordinary matter collapsed into the galaxy at the center of figure 3 by losing energy. It lost energy to radiation among other things. It lost energy to collisions. What kind of collisions ? Collisions having to do with electromagnetic forces. In the process of colliding, losing energy, radiating, it sort of fell into the center to some extent.

So what we need to expect then about the dark matter is that, whatever it is, it is much more weakly interacting. In particular it is not electrically charged.

Now of course if it were electrically charged, we would see it. It would radiate. So it is not electrically charged.

The fact that it is not electrically charged, and that it is very likely rather weakly interacting, maybe like neutrinos maybe a little stronger than that, means that the particles that make up the halo circulate around, go back and forth. Some of them go this way. Some of them go that way. Some have elliptical orbits. But they don't interact with each other very much.

Because they don't interact with each other very much, over the period of 10 billion years, the halo of dark matter has not collapsed and followed the luminous matter.

There is no question that to some degree dark matter and ordinary luminous matter do follow each other. It seems that there exists no large lumps of dark matter that don't have some kind of luminous matter at the center. And it seems that there are no galaxies out there, that are visible, that don't have dark matter around them.

That doesn't imply however that they behave the same in the detail. For instance the shape of the dark matter does not look like the somewhat flat disc shape of the galaxy. It looks like a great big 3D spherical halo.

Questions / answers session

Question : Did the first evidence of the probable existence of dark matter come only from looking at the speed of stars in the outer region of galaxies ?

Answer : No. Oort was looking at stars in the outer region of the Milky Way. But Zwicky was looking at the motion of galaxies in clusters of galaxies. So there was evidence of the phenomenon on multiple scales. It is not all coming from stars in one galaxy. It is also coming from interaction between galaxies. It is even coming from whole galactic clusters.

Since then many other people looked at dark matter. For instance Vera Rubin [12] attached her name to dark matter by studying in depth the motion of stars in galaxies to measure it.

We now also have plenty of evidence of dark matter from the deviation of light coming from behind with respect to us. It is a bit like we can say that a transparent drinking glass is full of water or not by the kind of distorsion of images it creates.

Q. : Does dark matter interact gravitationally like ordinary matter ? What do we know about its composition ?

A. : Everybody believes dark matter is made out of particles. They interact weakly. And we don't know yet what they look like. But they do not lead us to question Newton's laws. They gravitate just like any other massive [13] particles.

Since they are not electrically charged, they do not interact with electrical phenomena. And they do not radiate elec-

12. Vera Cooper Rubin (1928 - 2016), American astronomer.

13. In physics, massive does not mean huge, but made of mass, as opposed to made of pure radiation.

tromagnetic waves.

Furthermore the emission of gravitational waves by stars and other things moving in galactic orbits is completely negligible on the energy balance. So radiation of gravitational waves [14] is not an efficient way for things to lose their energy.

Q. : What do we know about the origin of dark matter and galaxies ? Also, could black holes appear directly from dark matter, or dark matter from black holes ?

A. : This was controversial for a while. I don't think it is controversial anymore. The idea is that dark matter appeared first. The galaxies formed by first the dark matter collecting in these great big haloes, then the baryons, the protons, neutrons, electrons and so forth, falling into the dark matter haloes.

Then the smaller galaxies formed. Stars formed. I don't know how early the black holes formed. And I don't know if anybody does. The black holes were probably simply the consequences of a lot of stars falling together.

Aside from the similar adjectives, black holes and dark matter have nothing to do with each other. Black holes are ordinary matter which collapsed and created singularities in

14. To learn more about gravitational waves see chapter 10 of volume 4 of the collection *The Theoretical Minimum*, on general relativity. Gravitational waves have finally been detected by observation in 2015 with the LIGO experiment. The results were published in February 2016.

spacetime as a result of extreme concentration. It is not that the dark matter formed around the black holes. And anyway, as mentioned earlier, black holes are a negligible amount of the mass of the center of galaxies.

So the galaxies formed as a consequence of the haloes. Stars formed. Some black holes appeared. That is probably the way things happened.

Q. : Is the density of dark matter in the halo homogeneous or does it vary with r ?

A. : It varies with r. We think that the density of dark matter is significantly higher near the center of the halo – and therefore of the galaxy – than away from the center.

For example, people who are looking for radiation coming from the annihilation of dark matter – two dark matter particles colliding, annihilating, and forming radiation – would look toward the center of the galaxy.

Q. : Is there any way we can know at what level dark matter is aggregated ? Is it gas, or cloud of dust, or what ?

A. : We don't know. But it is probably just lonesome particles. What does dust mean ? It means assemblages of large numbers of particles that hold themselves together. Together by what ? By electrostatic forces, by all the usual things that hold things together. If those kind of forces existed between those particles, they would not have survived as these kind of dark matter particles. So the expectation is that they are simply lonely particles.

Q. : If the Sun is moving in space through these particles, wouldn't they tend to be trapped in the solar gravity well ?

A. : Yes there is probably some abundance of them that is somewhat trapped near the Sun.

Q. : What explains the spiral arms of the galaxies ?

Figure 4 : Spiral arms of a typical galaxy.

A. : There are many theories about the spiral arms of galaxies. You should go look them up, and see for yourself which one convinces you.

Some people think that they are shock waves of star formation, shock waves propagating around in circles and not moving with the galaxy, nor with the dark matter.

There are other models, none of which I understand.

Q. : Could the dark matter be made of neutrinos ?

A. : Neutrinos are very light. As a consequence they would have to be pretty relativistic – i.e. go very near the speed of light – to have that much mass.

One fact about the particles of dark matter, which is most important, is that they tend to cluster. The gravitation makes them cluster.

Now things cluster less when they have high velocity. If these particles had high velocity they would tend to cluster less than what we observe.

Obviously if you have a gas of stuff, and it has gravitational attraction, and it is hot, it will resist the clustering tendency due to attractive forces. If it is cold on the other hand, it will tend to cluster.

Neutrinos being so light will tend to be pretty relativistic. They would be what is called *hot dark matter*.

But hot dark matter is a hypothesis that has been abandoned. Basically because of this tendency to cluster, dark matter must be cold.

Cold means that the particles are moving relative to the local reference frame with much less than the speed of light.

So neutrinos are not a good option for the dark matter.

Q. : In equation (7), with the four parts contributing to H^2, you said that the curvature part dominates over the matter part. But then you also said that, in a matter-dominated world, a goes like to $t^{2/3}$. So I'm a bit confused.

A. : a goes like to $t^{2/3}$ only up until the time when $|k|/a^2$ gets bigger than c_M/a^3.

The equation that says that $H \sim 1/T$ would have to be somewhat corrected, but not a lot because most of it was happening earlier, before the crossover point. And you can work it out.

Q. : Today what is dominating, the curvature part or the matter part ?

A. : Today according to yesterday's theory of fifty years ago, or today according to today's theory ?

Let's continue with the theory as it went fifty years ago. There was massive confusion.

If you put things together as they did, without worrying about dark matter, and also, as we saw, considering Λ to be zero, you would have had to say that the curvature term was the bigger in absolute value, that k was negative, and that a was roughly the same as the age of the universe – as we explained above in equation (10) and after.

People resisted it to some extent however. In truth each person had his or her own view.

But one of the prejudices, at the time, was that most people were convinced that the universe was closed and bounded. So they would tend to write equations that were appropriate for k positive. And then it conflicted with the observations. So there was a lot of confusion about it.

Later on, about twenty years ago, it sort of settled down. The dominant view became that k was probably negative.

Now, let me explain what the dark matter does to c_M in this equation

$$\frac{c_R}{a^4} + \frac{c_M}{a^3} + \Lambda - \frac{k}{a^2} = H^2$$

If we just ignore the curvature term $-k/a^2$, and the cosmological constant Λ is taken to be zero, the matter part on the right and the Hubble constant on the left differed by a factor of 25 approximately [15]. So this was a serious gap between the right-hand side and the left-hand side of Friedmann equation.

At best the dark matter was five or six or seven times the luminous matter, and it did not fix the problem. It didn't make the matter term equal to the Hubble constant term. What it did was : not only it left room for a k/a^2 term,

15. There was no question that the radiation part c_R/a^4 has become negligible after the very early history of the universe. And this view is not questioned today either.

with negative curvature, but it required it.

When I say it did not fix things, I'm not sure what fix means. There was nothing particularly wrong with saying that H^2 largely came from the curvature term. It was observed simply that the matter term alone was not sufficient to fill the gap.

Q. : Could not the nature of the dark matter particles solve the problem, for example if they were closer to light than to matter ?

A. : It depends. If the particles are fermions with very light mass you are in trouble. They can't all occupy the same state and be moving together slowly relative to each other. If they are bosons they can be extremely light and be in a phase called a Bose condensation [16]. A Bose condensation just has very slow particles all in the same quantum state.

Speculation that dark matter could be a Bose condensation is one of the theories about it, called the *axion theory* [17].

In this theory, the axions are incredibly light particles, much lighter than neutrinos. And they form a Bose condensation. It is a good theory. It is not a crazy theory.

But mostly people think about theories in which the dark matter is made of particles that could be detected in big accelerators – particles whose mass are in the range of masses

16. Also known as a Bose-Einstein condensate.
17. Proposed by Roberto Peccei and Helen Quinn in 1977.

that could be discovered at the LHC [18]. Is that wishful thinking ? It is anybody's guess.

Q. : Couldn't the detection of dark matter particles be beyond the theoretical maximum detection capability of the LHC ?

A. : No. There is no theoretical maximum, unless you assume that the particles interact with roughly standard model weak interaction cross-sections. Then the maximum is probably up in the range of a few TeV [19] [20].

So, was there enough energy to detect them within the standard theory ? As of March 2017, apparently not.

Q. : Would it be fair to say that dark matter – and string theory would bear the blame for it – is in other dimensions than the three spatial dimensions we are familiar with, and that would explain why we cannot see it ?

A. : No. Not at all. They are particles that can be seen.

18. Large Hadron Collider, particle accelerator located at the CERN in Geneva, Switzerland, used to create particle collisions allowing physicists to test the predictions of different theories of particle physics.

19. Tera-electron-volt. One TeV is about the kinetic energy of a flying mosquito.

20. It is long since physics observes new features of nature only indirectly – if direct observation means anything. Therefore there is no limitation to detection. Detection means effect. So does existence. Although, as said before, we can make sense of observations and decide that something exists, i.e. has been detected, only within the framework of a theory.

They can be detected and measured. In fact, we do measure them. We measure them gravitationally. For instance they create gravitational lenses.

I guess what you are asking is : is it conceivable that these particles are so unusual that they have almost no interaction with ordinary material, and that would be the reason why we don't see them ? Yes, that is possible.

It is possible that they are so weakly interactive with ordinary matter that there is no chance of detecting them with conventional experiments such as carried out at the LHC.

But the idea that we could observe things when they are in the usual 3D, but not when they are hidden in other dimensions, is nonsensical. We detect things when they create effects we can observe. The space in which they "are" is any mathematical space your whim has constructed for the theory.

For instance, alternative currents are represented with complex numbers. Can we observe them ? Yes. Are complex numbers in 3D ? It is a meaningless question. What is meaningful is that the representation of alternative currents is in the set of complex valued functions of time. That is not 3D.

In physics, however, as well as in daily life, whatever experiment we carry out is carried out in 3D, *in the sense that* it is the usual representation of the world in which we live and use tools. Experiments are in 3D ; effects we observe and measure are in 3D. That doesn't mean that the concepts we handle through experiments are necessarily in

3D. In the end, it all depends what you mean by to be somewhere.

Q. : Neutrinos are so fast, we are told, because they were formed during the Big Bang. Does this mean that the particles of dark matter were formed later when the universe was cooler ?

A. : No. No. They are just heavier. If they are bosons, it is okay : they can Bose condense and they don't have to be moving fast. If they are fermions, they have to be moving fast.

The axion theory is not a theory in which the axions were ever in thermal equilibrium. If they were in thermal equilibrium at one time, then they would be moving fast today. But this is not where I intended to go with this lesson.

We know with any confidence nothing about dark matter. The most likely hypothesis is that it is some form of elementary particle that hasn't been discovered yet. There is a good chance that they are light enough to have escaped detection in the LHC, but heavy enough to be able to be detected in future experiments. But that is it. That is about all we can say at the moment. And they are there.

The other alternative to dark matter was to modify gravitational theory so that Kepler's law doesn't work with galaxies. But that doesn't seem to have worked very well. There doesn't seem to be any nice modifications of standard gravity theory which explain Oort's and Zwicky's and other's observations. So most of us think it is particles. But you are free to make other theories.

Q. : We haven't had to introduce modifications to the Newtonian equations by which we understand our solar system [21]. Can we conclude from that that there is a negligible amount of dark matter around us ?

A. : That is correct. Keep in mind that the density of matter in the solar system is a lot higher than it is out there. The dark matter in the solar system does not compete with the gravity of the Sun at the center, not by a long shot.

Let's end here the session of questions and answers.

It is natural now to ask how do we test the candidate theory of the history and the present state of the universe ?

Testing the theory of the universe

We still place ourselves in the position of physicists a few decades ago. The candidate theory that we want to test is that the universe at present is matter-dominated. So we start again from the fundamental Friedmann equation

$$\frac{c_R}{a^4} + \frac{c_M}{a^3} + \Lambda - \frac{k}{a^2} = H^2 \tag{26}$$

21. Like did for instance the French astronomer and mathematician Urbain Le Verrier (1811 - 1877). He did not modify $F = m\ddot{x}$, but introduced a new hypothetical planet to try and explain the discrepancies in the trajectory of the planet Uranus from what it should have been. And this lead him to the discovery of the planet Neptune.

We forget about c_R/a^4 because we know it is negligible. We forget about Λ because it is something we haven't heard about. So equation (26) reduces to

$$\frac{c_M}{a^3} - \frac{k}{a^2} = H^2 \tag{27}$$

The curvature parameter k is negative. Otherwise equation (27) would be impossible because H^2 is way too big compared to c_M/a^3. So we are in a *negatively curved*, or said another way *hyperbolic*, universe.

The reader remembers that its full spacetime metric is given by

$$ds^2 = -dt^2 + a(t)^2 \; [\; dr^2 + \xi^2(r)\; d\Omega_2^2 \;] \tag{28}$$

where $a(t)^2$ times the expression in square brackets is the spatial part of the metric, $a(t)$ is the scale parameter, and $\xi(r) = \sinh(r)$.

How do we test that theory? In fact how do we test any particular model?

A model consists of two things :

1. A specification of k, whether it is 0, $+1$ or -1.
2. Some kind of equation of state.

For the equation of state we can substitute instead a history of the scale factor, i.e. a function $a(t)$. In other words, the choice of an $a(t)$ and a k defines a model that we can try to test.

So what do we know? What can astronomers measure? And how do they compare it with a particular model?

We do pretty much the same thing Hubble did, except a more sophisticated version of it. We can measure essentially two things. The redshift of astronomical objects tells us something about their velocity. And we can measure distance by luminosity. How do those two things come together to tell us anything about about k and $a(t)$?

First let's ignore the fact that $a(t)$ changes with time. In other words, let's just suppose we didn't have to worry about the fact that the universe evolves. We suppose for a minute that it was all at an instant and it didn't change. What could we do to measure k, to find out if the universe is closed and spherical, or infinite and flat, or else infinite and hyperbolic?

We have already talked about the procedure in chapter 3, but we shall go over it again to insist on its practical relevance. To figure out the curvature of the universe, given that, by hypothesis, it is homogeneous, we measure the number of galaxies out there of a given brightness.

What does the brightness have to do with k? As said, measuring brightness, in other words luminosity, is measuring distance. So we shall be counting the number of galaxies as a function of distance really.

Mind you, if you happen to think that distance is related to redshift – which of course it is – then you could also be talking about the number of galaxies as a function of redshift. But let's just talk about the number of galaxies as a function of distance, *where the distance has been measured in some way that we will need to come back to.*

If we live in a flat 3D space, and as we stressed it is a homogeneous space, then it is pretty clear the number of galaxies in a spherical shell between radii r and $r + dr$ goes like $r^2\ dr$, figure 5.

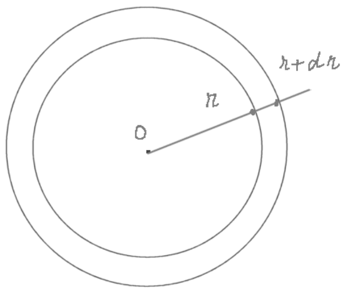

Figure 5 : Shell between radii r and $r + dr$. Imagine the picture in 3D, and the circles as spheres.

This is because, in flat 3D space, $r^2\ dr$ is proportional to the volume of the shell.

It can be generalised to hyperbolic geometry : if we live in a negatively curved space, the number of galaxies in a shell between radii r and $r + dr$ would go like $\sinh^2 r\ dr$. So it would grow exponentially with r since, for large r, hyperbolic sine of r is approximately equal to one half exponential of r. Thus, paying attention to the square of sinh, if $k = -1$ the number of galaxies between distances r and $r + dr$ goes like $e^{2r} dr$.

If the universe was a 3-sphere, then the number would grow like $\sin^2 r\ dr$. But let's leave aside the spherical case, because the evidence says that in equation (27) k is equal to -1, and that we are in the hyperbolic case.

So the first thing we see is that the number of galaxies, between r and $r + dr$, as a function of distance would grow much more rapidly in hyperbolic space than in flat space.

But remember that we use luminosity [22] to figure out distance. So what can we say about the number of galaxies between r and $r + dr$ as a function of luminosity ? Another question we could ask is : how does luminosity depend on distance ?

In flat three-dimensional space, luminosity goes like $1/r^2$. It is just the same fact as illustrated in figure 5. If we thought of a light bulb at the center O, and us out at distance r, the light emitted by the bulb would simply be spread over the area of the sphere where we are [23]. And the area grows like r^2. So the luminosity itself goes like $1/r^2$.

But supposing we were living in a negatively curved three-dimensional space, where the area of the sphere at radius r grows like e^{2r}, then the luminosity of an object at radius r would be much smaller than in flat space. We would find, as the distance increases, a very rapid increase of the number of galaxies between r and $r + dr$, as well as, inversely, a very rapid fall off in luminosity.

22. Luminosity at distance r is defined as *light per unit area* received at that distance.

23. The *total* amount of radiation is the light emitted by the bulb. That does not depend on distance. But the luminosity perceived at r does depend on the radius r.

Exercise 2 : Suppose we have a calibrated candle such that at distance 1 its luminosity is 1. Then suppose that we see a similar candle with luminosity $1/400$. At what distance is it ? Show that

a) **if we know we are in a flat space, then it is at distance 20,**

b) **whereas if we know we are in a hyperbolic space of radius 1, it is only at distance 3.**

We could also ask : what about luminosity as a function of redshift ? We would also find that in a hyperbolic space luminosity gets small really fast, again because of the factor e^{2r}.

That is pretty much the way of determining whether k is $+1$, -1 or 0. In particular, if $k = -1$, meaning to say a hyperbolic space, one should find that the number of galaxies at a given distance, or let's say at a given redshift, grows anomalously rapidly with redshift. As the redshift gets deeper into the red, the number of galaxies should grow much faster than it would grow in flat space. So that is a test.

You can also think about calculating the number of galaxies that you see as a function of redshift. Forget the dimness and the brightness for a minute now. Just how many galaxies do we see at each redshift, or more precisely between each increment of redshift ? That is something we can also calculate. Let's talk a little bit about how we do that.

In order to talk about that kind of measurement and that kind of test, we have to talk about redshift.

Redshift

If light is emitted by a source that has a wavelength lambda, we can call that *the wavelength of the light when it is emitted*, and denote it

$$\lambda_{emitted}$$

It is emitted the usual way. It is simply a light bulb that has a certain wavelength.

When it is detected, the light may or may not have the same wavelength. There could be various reasons for a shift of the wavelength. Doppler shift is one. Gravitational redshift from gravity due to a black hole for instance is another one. The expansion of the universe is also a possible reason, although it happens to be equivalent to the Doppler shift. So it is not a third one. However it is the one we are interested in.

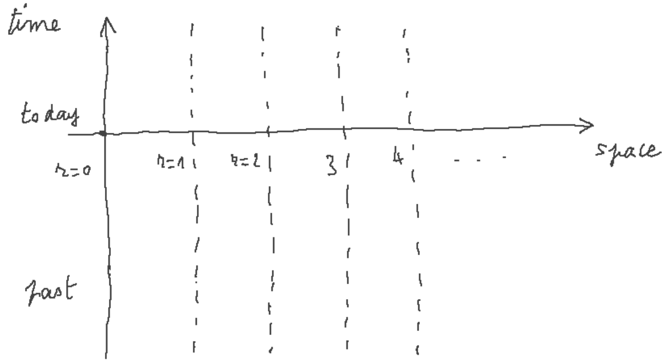

Figure 6 : Spacetime. For convenience, as usual, space is represented as one-dimensional.

Here is what is going on. Let's draw a map of spacetime, figure 6. The vertical axis is time. The horizontal axis is space. The plain horizontal line corresponds to today. We are at the point $t = 0$, i.e. today, and $r = 0$. And we drew vertical lines corresponding to various values of r, that is, distances from us.

We look at a light ray arriving at us now. How do light rays move ? In flat space, in the Minkowski metric, they move on 45° lines.

But this is general relativity. There is a complicated metric here. We will discuss a little later how a light ray moves. Let's represent one, however it has been moving, either a straight line or a curved trajectory, figure 7.

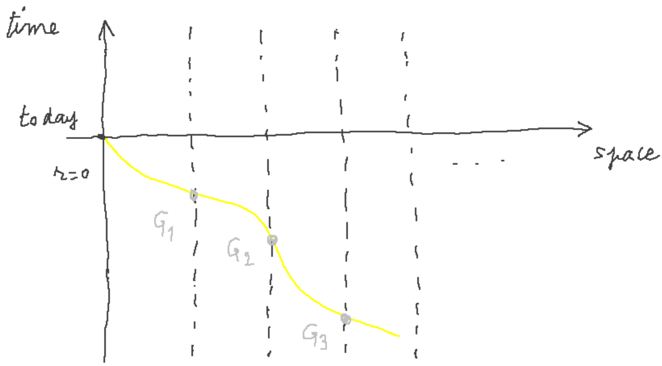

Figure 7 : Trajectory of a light ray reaching us from the past.

The light ray reaching us at time $t = 0$ carries light coming from galaxy G_1, which was at radius $r = 1$ at some time t_1 in the past, from galaxy G_2, which was at radius $r = 2$ as some time t_2 in the past, from galaxy G_3, and so forth.

We see the different galaxies as they were at different times. And, because we see them at different times, we see them first of all with different scale factors. The scale factor $a(t)$ varies vertically in figure 7. For each horizontal line there is a t and therefore an $a(t)$.

Let's focus for instance on galaxy G_3 that was at position $r = 3$ at time t_3 in the past. For clarity, let's call that time the *emission time* t_e. In other words, $t_e = t_3$.

The light coming from G_3, that is, that was emitted at the point in spacetime[24] $(r = 3,\ t = t_e)$, has to stretch to get to us. Indeed between t_e and today, which is $t = 0$, the universe expanded by a factor

$$\frac{a(0)}{a(t_e)} \tag{29}$$

The value $a(0)$, of the scale parameter a today, can just be thought of as a fixed number. On the other hand, the value $a(t_e)$, that is, the value of the scale parameter at the emission time, depends on that emission time.

$a(0)/a(t_e)$ is the factor by which the wavelength of an electromagnetic wave will stretch in going from G_3 to us in figure 7. And it is simply related to the *redshift factor*.

Let's define the redshift factor precisely : it is the ratio of lambda when it is detected divided by lambda emitted, minus one. It is called Z.

24. In special relativity, we call these points *events*, see volume 3 in the collection *The Theoretical Minimum*, on special relativity and electrodynamics.

$$Z = \text{redshift factor} = \frac{\lambda_{detected}}{\lambda_{emitted}} - 1 \tag{30}$$

Why is there a minus one in the definition of Z? For historical reasons. The important physical fact is that

$$\frac{\lambda_{detected}}{\lambda_{emitted}} = \frac{a(0)}{a(t_e)} \tag{31}$$

For example, if the scale factor $a(t)$ went from 1 to 1.2 between t_e and now, the redshift factor for that light is 20%.

So when we look at a light ray arriving at us at time 0, it carries light from different events in the past. We see light that was emitted from G_1 at time t_1, we see light emitted from G_2 at time t_2, we see light emitted from G_3 at time t_3, etc. That is why looking at a starry night sky is also looking at the past history of the universe, figure 8.

Figure 8 : Looking at a starry night sky is also looking at the past history of the universe.

Back in our spacetime with only one spatial dimension, as in figure 7, all the lights, received together today in the light ray arriving at us, were emitted at different times in the past. And *they display different redshift factors.*

Since from equations (30) and (31) we have

$$Z(t_e) = \frac{a(0)}{a(t_e)} - 1 \tag{32}$$

if we know the value of the scale factor today, and at the time of emission, then we know what $Z(t_e)$ is.

But so far we don't know anything, because we don't know how the light ray moves in figure 7. Until we know how the light ray moves, we don't know how to compare the points G_1, G_2, G_3 ...

So let's talk about how the light ray moves. Let's workout all the things we need, and then we will write down some equations, which an astronomer would actually work with.

In order to figure out how a light ray moves, we go back to the metric.

$$ds^2 = -dt^2 + a(t)^2 \; [\; dr^2 + \xi^2(r) \; d\Omega_2^2 \;] \tag{33}$$

Light rays are *null rays*, i.e. follow trajectories along which $ds = 0$.

Since in figure 7 the spatial part of spacetime is only one-dimensional, there is no term $\xi^2(r) \; d\Omega_2^2$ in equation (33). Therefore, for light rays moving to and from us, we simply have

$$dt^2 = a(t)^2 \; dr^2 \tag{34}$$

Obviously the one we are interested in is the one which goes into the past, or, said more properly, comes from the past. It satisfies

$$dt = -a(t) \; dr \tag{35}$$

This equation says that, as t gets more negative, r gets bigger. It can also be seen of course in figure 7. Or we can write equivalently

$$dr = -\frac{dt}{a(t)} \tag{36}$$

This is one equation. And it tells us how light moves along the path shown on figure 7.

Let's now invent a question that an astronomer would be interested in :

What is the number of visible galaxies of a given redshift ?

Of course we are really talking about a density. So the question is more accurately : what is the number dN of visible galaxies within redshift Z and $Z + dZ$? It is the derivative function

$$\frac{dN}{dZ} \tag{37}$$

That is a function of Z that surely can be measured for any value of the redshift. Look out and count the galaxies, or whatever the standard candles are you are using, in each slice of redshift. Let's see if we can figure out an equation for it.

First, what about dN ? Go back to figure 5. Let's take dN to be the number of galaxies within the two spheres of radii r and $r+dr$. The reasoning doesn't depend on the geometry of the space, that is on the value of k. Then we shall link r and Z.

What is that dN going to be ? It is going to be equal to the area of the sphere of radius r times dr. And what is the area of the sphere or radius r ? Is it $4\pi r^2$? Not in general. That would be correct in flat space. In general it is proportional to $\xi^2(r)$ with a proportionality factor $c(k)$. Let's denote that surface $c(k)\ \xi^2(r)$. So the derivative function (37) can be rewritten

$$\frac{dN}{dZ} = \frac{c(k)\ \xi^2(r)\ dr}{dZ} \text{ times a density} \tag{38}$$

In the case of interest, where $k = -1$, $\xi^2(r)$ would be $\sinh^2(r)$. And at large distances it would be some exponential. But of course we can study all of the cases simultaneously by keeping $\xi^2(r)$ in equation (38). Furthermore the unspecified density is the number of galaxy per unit volume in the universe. Remember that, as a consequence of the hypothesis of homogeneity, it is a fixed number. Rather than carry along various constants which don't play a role in the subsequent calculations, let's just write

$$\frac{dN}{dZ} \sim \frac{\xi^2(r)\ dr}{dZ} \tag{39}$$

where the sign $\sim$ means proportional to. On the numerator, the number of galaxies within a shell of radius r and width dr is proportional to $\xi^2(r)\ dr$.

Now, what about dZ ? Let's look at Z given by equation (32). We are interested in the change in Z when going from one radial point to another radial point at distance dr nearby, see figure 7. We shall just differentiate the expression of Z. The term -1 will disappear.

What is varying in equation (32) ? Is $a(0)$, also called a_{today}, varying ? No it is not : a_{today} is just a_{today}. On the other hand, $a(t)$ at emission time will vary when we move on the trajectory of the light ray. Applying the usual formula for the derivative of a function in the denominator, we can write

$$dZ = -\frac{a_{today}}{a(t)^2}\ da \tag{40}$$

Does the reader recognize this equation ? Let's plug-in this dZ into equation (39). We get

$$\frac{dN}{dZ} \sim \frac{\xi^2(r)\ dr}{-\frac{a_{today}}{a(t)^2}\ da} \tag{41}$$

What is da ? It is da/dt times dt. So we can also write

$$\frac{dN}{dZ} \sim \frac{\xi^2(r)\ dr}{-\frac{a_{today}}{a(t)^2}\ \frac{da}{dt}\ dt} \tag{42}$$

What is dt ? It is not quite dr. The scale factor plays a role. We have, from equation (35), $dt = -a(t)\ dr$. Using the expression for dr/dt, we can simplify equation (42). The minus sign disappears, and one of the factors $a(t)$ cancels. We get

$$\frac{dN}{dZ} \sim \frac{\xi^2(r)}{\frac{a_{today}}{a(t)}\ \frac{da}{dt}} \tag{43}$$

Let's replace $\frac{a_{today}}{a(t)}$ by $Z+1$. The equation becomes

$$\frac{dN}{dZ} \sim \frac{\xi^2(r)}{(Z+1)\ \frac{da}{dt}} \tag{44}$$

So, here we have the number of galaxies, seen at redshift Z, or more precisely their density there. It contains a factor $1/(Z+1)$. That is known if we know Z. And then there is $\xi^2(r)$ divided by $\dot{a}$. What can we do with it ? Nothing until we make a model. So let's make a model and see what we get.

Our model is the matter-dominated universe, with $k = -1$. In this model, the reader remembers, for most of the evolution, the scale factor satisfies

$$a(t) \sim t^{2/3} \tag{45}$$

Here is what we want to do. We won't go through the whole story in detail. First of all we calculate $\dot{a}$.

$$\dot{a} \sim \frac{2}{3}\ \frac{1}{t^{1/3}} \tag{46}$$

What about r ? Can we calculate it ? Yes, once we know $a(t)$, we can calculate r. We use again equations (35) or (36). Let's not go through the calculations, but $r(t)$ is known.

Moreover, once we know a, we can calculate Z.

The point is that everything, a, $\dot{a}$, r, can be expressed in terms of Z and a_{today}. We know the relationship between a and Z, equation (40). We know the relationship between a and t, equation (45). We know $\dot{a}$, etc. In other words, we can write everything in terms of Z.

On the right-hand side of equation (44) we get something which depends only on Z. But pay attention to the fact that it is known only because we substituted a particular model. Had we used a different equation of state [25], we would have got a different answer.

$\xi(r)$ is an interesting quantity. It exponentially increases with r. It is really easy to detect the explosive behavior with increasing r. Since r increases with redshift, it makes an anomalously large number of galaxies at large redshift if k is negative. And $\dot{a}$ can be computed directly in terms of Z.

In summary, with a little bit of labor, we can take a model – which means an $a(t)$ and a k – and convert it into a statement of how many galaxies we see per unit redshift, which is the meaning of dN/dZ. The other important piece of information we can get from the model is the relationship between redshift and luminosity, that is a function denoted $L(Z)$.

Conversely, the observation of those two bits of informa-

25. The equation of state is the fundamental thermodynamic equation for the universe which enables us to figure out the function $a(t)$. See the equation of state (23) of chapter 4. And in the present chapter we just used the model $a(t) \sim t^{2/3}$. For these reasons we sometimes slip into calling the formula for $a(t)$ the equation of state.

tion, dN/dZ and $L(Z)$, can give us a complete history of the expansion of the universe, and tell us what $\xi(r)$ is.

In the above calculation, we implicitely assumed that we knew $\xi(r)$ corresponded to the hyperbolic case. But we didn't have to. We could have done the calculation in flat space. We could have done it in a closed universe.

Since it is important, we repeat it : the two pieces of information, luminosity as a function of redshift

$$L(Z) \tag{47a}$$

and the number of galaxies as a function of redshift

$$\frac{dN}{dZ} \tag{47b}$$

are enough to determine $a(t)$ and k, whether it is $+1$, 0 or -1.

So we do that. Astronomers do that. They have been doing it for fifty years or more. And the final upshot is first of all inconsistent with the matter-dominated universe, given by equation (27), with k being negative.

Incidentally, when we talk about dN/dZ we don't necessarily talk about galaxies. We talk about standard candles of some kind, i.e. objects out there in the universe which have a calibrated behavior and enable us to measure distance. The standard candles which are best for the redshift we are interested in are supernovae. They are very controllable. We can't control them but nature controls them.

What are the observations consistent with ? The result is not equation (27) at all. Going back to equation (7), expressed with proportions, see equation (8), it is this :

a) $\Omega_R \approx 0$

b) $\Omega_M \approx .3$. This includes luminous matter and dark matter.

c) $\Omega_\Lambda \approx .7$

d) $\Omega_K \approx 0$

This is an observational fact. It is the values of the omegas, that is the proportions of the four types of energy, *today* in the history of the universe. Very little theory has gone into it. Well, some theory has gone into it, but no particular biases about what to expect. This is pretty much what the unbiased raw data say.

When we write $\Omega_K \approx 0$, in fact it could be up to $+0.01$, meaning a hyperbolic space almost flat (remember that k and Ω_K have opposite signs), or it could be down to -0.01, meaning a spherical space almost flat, like the surface of the Earth around us is spherical but almost flat. But it could not more in magnitude.

The four proportions above correspond to the best fit with the observed informations about (47a) and (47b), that is about luminosity and density of candles as a function of redshift, using models which depend on the equation of state.

So, just from counting galaxies, redshift and luminosity, this is the picture of the universe at time t_{today} which emerges.

Questions / answers session (2)

Q. : Is that what was believed a few decades ago, in the 1960's and 1970's ?

A. : No, no. We are talking now about what is the picture today.

Today has two meanings : it is the time t_{today} in the equations, be they those written in 1970 or in 2017 ; and it is the present days in the first quarter of the XXI$^{\text{st}}$ century.

The proportions given above is the picture of the universe today (at t_{today}) seen by scientists today (in 2017).

The theory failed, that postulated only

$$\frac{c_M}{a^3} - \frac{k}{a^2} = H^2$$

It just could not predict or reproduce by calculations, with any equation of state and k, the luminosity $L(Z)$ and density dN/dZ, as functions of Z, that we observe.

Q. : The way we know $\Omega_\Lambda \approx .7$ is only from the equations ? We are not measuring Ω_Λ, are we ?

A. : Right. Ω_Λ is not directly measured.

What we do is *we start with* a hypothetical set of numbers for Ω_R, Ω_M, Ω_Λ, and Ω_K.

Or equivalently we start with a hypothetical set of values for c_R, c_M, Λ and k.

Then we go through an iterative process of trial and error. One iteration is as follows :

1. From the hypothetical set of numbers, we build a model. That is, we write the Friedmann equation with actual values for the parameters. For the parameter k we try in turn -1, 0 and $+1$.
2. We solve Friedmann equation for $a(t)$.
3. From $a(t)$ and k, using the equations (35) to (44), we compute the theoretical $L(Z)$ and dN/dZ.
4. We compare them with observations.
5. If it fails we throw the model away and put in some new numbers.

The set of numbers $(0,\ .3,\ .7,\ 0)$ for $(\Omega_R,\ \Omega_M,\ \Omega_\Lambda,\ \Omega_K)$ works best. There are some uncertainties, but not big, of the order of 1%. If you vary the input numbers $(0,\ .3,\ .7,\ 0)$ very much, by more than 1%, the test fails.

Again, a model consists of a specification of all the parameters in the complete Friedmann equation[26] , that is equation (7). But we don't measure them all. Some of them we can measure. Some we can't measure, but we obtain as an output of the above iterative process.

It is true that c_R and c_M are measurable more or less directly. c_R can be neglected, because it played a role only in the very early history of the universe. Now, when divided by a^4, it plays no role. c_M, combining luminous and

26. Remember that H is not a parameter, it is the function $\dot{a}/a$.

dark matter, is such that Ω_M today is about 0.3. $H(t_{today})$ is measurable too, but anyway it is not a parameter to be adjusted. We get it as an output from the model.

The other two parameters, Λ and k, come only from working with the model, in the trial and error process, and looking for a best fit between theoretical predictions and observations of $L(Z)$ and dN/dZ.

The relation between the redshift Z and the luminosity is a generalization of Hubble's original calculation. The reason why we have to be more sophisticated than Hubble's first calculation is that we are looking at different times. And an additional ingredient is the possibly *curved* trajectory for the light ray in figure 7.

Q. : Where does a_{today} come from ? Is it from the Hubble constant H ?

A. : You could say that. But remember that H depends on time. It is better to understand that we are juggling here with various elements.

But once we have a model that fits the two pieces of observational data $L(Z)$ and dN/dZ, then we have a scale factor a as a function of time. We can read off what a_{today} is. We can also figure out the Hubble constant, $\dot{a}(t)/a(t)$, as a function of time, as well as its value today. All the parameters are linked.

The equation whose predictions we are trying to fit is Friedmann equation (7). Remembering that H is itself a function of $a(t)$, we can rewrite it in a way which may clarify the trial and error process just described.

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{c_R}{a^4} + \frac{c_M}{a^3} + \Lambda - \frac{k}{a^2} \tag{48}$$

The four parameters we want to adjust are c_R, c_M, Λ and k. Those are assumed to be independent of time. Equation (48) is a differential equation whose unknown is the function $a(t)$. A set of numerical values for these four parameters leads to a solution function $a(t)$. In practice all this is done with numerical methods on computers.

However we solved equation (48) in closed-form in many cases in the preceding chapters. For instance, we solved

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{c_M}{a^3} \tag{49}$$

We solved

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{C}{a^2} \tag{50}$$

We solved

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{c_R}{a^4} \tag{51}$$

We solved

$$\left(\frac{\dot{a}}{a}\right)^2 = \Lambda - \frac{k}{a^2} \tag{52}$$

And we solved several more.

We invite the reader to review the various models of universe we investigated in chapters 1 to 5, and each time the equation it lead us to. The reader should also review the important notion of *equation of state*, introduced in chapter 4, which produced different possible right-hand sides for equation (48) via an energy density ρ.

Thus solving Friedmann equation yields a function $a(t)$. Then $a(t)$ leads to theoretical functions for $L(Z)$ and dN/dZ. We can check them against observations. That is how we find the best parameters c_R, c_M, Λ and k, and consequently the best values of Ω_R, Ω_M, Ω_Λ, and Ω_K, which all depend on time since a and H depend on time.

Let's stress that once we have obtained the best parameters c_R, c_M, Λ and k, these don't change with time. So looking again at Friedmann equation, for instance in the form

$$\frac{c_R}{a^4} + \frac{c_M}{a^3} + \Lambda - \frac{k}{a^2} = H^2 \tag{53}$$

we see that long in the past, in the early part of the history of the universe, the dominant term was c_R/a^4. In other words, a long time ago Ω_R was large – meaning it was almost 1, and the other omegas were essentially 0.

Long in the future Ω_Λ will be 1, and everything else will be small.

The values of Ω_R, Ω_M, Ω_Λ, and Ω_K, at time t_{today}, are estimated to be respectively 0, .3, .7, and 0.

These figures are all within 1%. For instance the last figure, being $-k$ divided by $a_{today}^2\ H_{today}^2$, might be actually 0 if

$k = 0$ and the universe is flat. But it has to be slightly positive, if the universe is hyperbolic with a very large radius today.

The next round of experiments might give more precise estimates and confirm in particular that k is negative and that Ω_K is, say, 1%. But the figures we just gave, for Ω_R, Ω_M, Ω_Λ, and Ω_K at time t_{today}, won't change much.

Of course, as time passes and $a(t)$ gets bigger, the terms divided by a power of a in equation (53) will get smaller. Eventually all there will remain is Λ. And that will correspond to a pure exponential growth, as we have already seen in a de Sitter space.

Q. : Could we figure out if k is -1, 0 or $+1$ by doing huge triangulations, computing the sum of the angles in trianges, the volume of tetrahedra and things like that in the universe ?

A. : Yes, that is correct.

We shall come to that in the next lesson when we study the cosmic microwave background (CMB). The CMB or at least the lumpiness of the CMB allows triangulation.

Using only supernova data, like $L(Z)$ and dN/dZ, doesn't get you to one percent precision in the estimation of the parameters. Maybe it gets you somewhere between ten percent and one percent. It is combined and supplemented with CMB data that we get to one percent precision.

Q. : As Ω_K gets closer and closer to zero, does this mean the universe is getting flatter and flatter ?

A. : Oh yes. That is true.

Remember that if the universe is not flat, $a(t)$ is the radius of the universe, or proportional to it. This is true for a 3D spherical universe, where the radius is somewhat intuitive [27]. And it is true for a 3D hyperbolic universe where there is also a notion of radius.

The curvature of the universe is the inverse of its radius. So the larger the radius, the smaller the curvature, and the flatter the space is.

Big is flat like, when looked close, smooth is straight,

meaning that when looked at with a magnifier any differentiable curve is almost like its tangent.

What we have learned in this lesson can be seen as tricky stuff. We have to familiarize ourselves with 3D spaces that can be flat, spherical or hyperbolic, and furthermore are expanding. We have to become accustomed to how distances, surfaces, volumes, real and apparent sizes and densities, be-

27. Don't mix up, though, a 2-sphere in expansion and its growing radius which we visualise easily, and a 3-sphere in expansion and its growing radius which is not as geometrically intuitive.

have in these spaces depending on curvature.

The only concrete example we can once in a while return to, in order to sustain geometric intuition, is how we live in a world that was for a long time thought to be flat, until some people in Antiquity began to suspect that it was more likely to be a huge sphere[28].

The process to estimate the proportions of the various types of energy today in the universe is also somewhat tricky.

Remember that we start with a model, which means a set of parameters in Friedmann equation. We solve it for $a(t)$. From there we make theoretical predictions about $L(Z)$ and dN/dZ. We compare them with observations. And, after a process of trial and error, we keep the parameters for which the fit is best.

There is, incidentally, no reason why any particular set of numbers has to agree with the data, other than theory.

So when we go through the process of trial and error to adjust the parameters, we are not only testing a particular model. We are also testing the theory.

28. But this example is still entirely perceived by us, in our mind, embedded in the usual 3D Euclidean space. We have to develop an intuition disconnected from such embedding, like the time-honored little bug in its 2D variety.