ASTR 541: Classical Cosmology
Jan 2024

In the next two weeks, we will cover the basic classic cosmology. The material is covered in Longair Chap 5 - 8. We will start with discussions on the first basic assumptions of cosmology, that our universe obeys **cosmological principles** and is expanding. We will introduce the **R-W metric**, which describes how to **measure** distance in cosmology, and from there discuss the meaning of measurements in cosmology, such as redshift, size, distance, look-back time, etc.

Then we will introduce the second basic assumption in cosmology, that the gravity in the universe is described by Einstein's GR. We will not discuss GR in any detail in this class. Instead, we will use a simple analogy to introduce the basic dynamical equation in cosmology, the **Friedmann equations**. We will look at the solutions of Friedmann equations, which will lead us to the definition of our basic cosmological parameters, the density parameters, Hubble constant, etc., and how are they related to each other.

1 Cosmological Principles

The crucial principles guiding cosmology theory are homogeneity and expansion.

The cosmological principle says: We are not located at any special location in the Universe. The other way to put it is that the universe is homogeneous and isotropic.

Isotropy. This is not so obvious in the nearby galaxies, but distant galaxies, radio sources, and the CMB all show isotropy.

We could be living at the center of a spherically symmetric universe with a radial profile. However, we probably are not at a special point. If two observers both think that the universe is isotropic, then it is in fact homogeneous.

Homogeneity is a stronger result, because it means that we can apply our physical laws from Earth. Homogeneity is harder to prove. But if we have a 3-D map of the sky, as in SDSS, we can certainly ask the question of how uniform (not isotropic) the universe is on a fairly large scale, which proves homogeneity. In addition, we see high-z object spectra, which follow the same quantum physics law; motions in distant clusters of galaxies follow the same gravity laws. Also, people can measure the CMB temperature at high-z, as the excitation temperature in low-T environment, which follows (1+z).

At low-redshift, this can be done by using so-called Sunyaev-Zeldovich effect in cluster of galaxies. In this case, the CMB photons are upscatttered through inverse Compton effect when encountering hot elections in the cluster plasma.

At mid-redshift, this can be done by measuring the excitation temperature of cold gas traced by quasar absorption line.

And a new work by Riechers et al., they modeled the exciting of a water line in z=6.3 star burst galaxy to constrain CMB temperature. All show that CMB temperature increases as (1+z).

Both of these ideas only apply on large scales, hundreds of Mpc. As we will see, the universe is several Gpc across, so there are many patches in the universe with which to define homogeneity. **scale: few hundred Mpc**

CMB plots.

For the next while, we will consider the universe to be straightly isotropic and homogenous to study the basic principles of cosmology. We will then try to relax this, since, as you know, if the universe is really 100% homogenous, then there will be no gravitational fluctuation and no grow of fluctuation, and therefore no formation of galaxies etc. So in later part, we will consider how the universe is different from completely homogenous, and how these fluctuations grow with time, from the very small fluctuations in CMB to forming large galaxies and large scale structure of galaxies we see today.

Before we move on, for completeness, we should discuss a stronger version of cosmological principles, i.e., the so-called **perfect CP**, of Bondi and Gold, and Hoyle in 1940s. This requires the invariance not only under rotation and displacements in space, but also as a function of time. In other words, the universe looks the same at all directions, all positions and all times. This hypothesis led to the **steady state cosmology**, which at least until the discovery of CMB in 1965, was a valid, competing, and some might say, sometimes dominating cosmology. However, people did know at that time that the universe was expanding. To solve that, steady state cosmology requires that the universe is in a continuous creation of matter to keep the mean density of the universe constant. You might think that this is odd; but so is dark energy or dark matter, so I don't think people was scared about this unknown physics. Even after the discovery of CMB, steady state cosmology didn't die easily and its proponents invented ways to explain CMB. It was really only after COBE etc., with the smoothness of the CMB, which is very difficult to explain without an extremely optically thick phase of the universe at high-z, that steady state cosmology finally left any main stream literature, and we should not discuss it again in this class.

2 Metric

The next key concept of modern cosmology is **metric, which specifies the distance between two points, in space or in 4-d space time**. Our goals here are:

1. how to express distance as a function of difference in coordinates, $\delta l(\delta x, \delta y, \delta z)$
2. how to incorporate time in our expressions and
3. how to incorporate expansion in the expressions.

In 1935, Howard Robertson and Arthur Walker derived (independently) the form of the metric of space-time for all isotropic, homogeneous, uniformly expanding model of the Universe. The form of the metric is independent of the assumption that the large-scale dynamics of the universe are described by GR. Whatever form of the physics of expansion, the space-time metric must be of RW form, because, as we will see now, the only assumptions are isotropy and homogeneity. Now let's see how to derive it.

Distant galaxies are observed to be redshifted, as if they are moving away from us. The recessional velocity is proportional to the distance of the galaxy. This is known as the Hubble law. This redshifting is easy to understand in a homogeneous universe, as we will see on Wednesday. All observers would agree on the expansion and the Hubble constant. However, the fact that the universe used to be smaller does cause some computational complications in the coordinate system. We won't be deriving cosmology from general relativity in this course, but we do need to use these coordinate systems.

Again, the metric specifies the distance between two points (in space or space-time).

Imagine a 2-dim flat space. We are used to measuring the distance between 2 points by the Pythagorean theorem.

$$(\Delta\ell)^2 = (\Delta x)^2 + (\Delta y)^2$$

If the two points are infinitesimally close together, then we can write the distance as

$$d\ell^2 = dx^2 + dy^2$$

However, the idea of differencing the coordinates only works in Cartesian coordinates. We are astronomers. We measure the position of astronomical objects using radial distance and angles, RA and Dec, as in polar coordinates. If we switch to polar coordinates, then the distance between two neighboring points is

$$d\ell^2 = dr^2 + r^2 d\theta^2$$

The formula for measuring the distance is called the metric. In general, there could be mixed terms, e.g. $dr\,d\theta$, but we won't need these much. Sometimes, one writes this as a matrix $d\ell^2 = g_{\mu\nu}dx^\mu dx^\nu$, in which case the matrix $g_{\mu\nu}$ is called the metric tensor. Note that the metric can depend on the coordinates themselves.

To compute the length of a curve in our space, we integrate the length $d\ell$ along the curve.

Next, imagine the surface of a sphere of radius R_c. We are interested in homogenous space, following the cosmological principles. Thee flat plane is an obvious example of a space in which all points are the same. The sphere is another example. use latitude and longitude θ, ϕ.

see draw figure on slides.

Here, using spherical coords, we have

$$d\ell^2 = R_c^2 d\theta^2 + R_c^2 \sin^2\theta \; d\phi^2$$

Note

$$dr = R_c d\theta,$$

then:

$$d\ell^2 = dr^2 + R_c^2 \sin^2\theta \; d\phi^2$$

The last expression is to use the polar coordinate on the surface of the sphere.

This differs from our result with polar coordinates on the flat space. One of the important results from differential geometry is that no change of coordinates can turn one metric into the other. In other words, there is no coordinate system on the sphere that looks like the Cartesian metric. This is because the sphere has a curvature that cannot be done away with by a coordinate transformation. Indeed, this leads one to the result that the metric gives a space its curvature without any reference to the embedding space of the manifold. Specifying the distances between all neighboring points defines the curvature of a space. Therefore, by measuring distances, we can derive the curvature of our universe. And if you believe in GR, that curvature of the space is caused by gravity of matter in it, then, measuring distance will measure the matter content of the universe, a topic we will discuss next week.

However, despite the apparent complexity, near any non-singular point, one can always define a coordinate system that appears Cartesian. In our case, near the North Pole, if $r \ll R_c$ we will recover the polar coordinate metric, which we know is just a transformation of the Cartesian one.

Our spherical metric yields circles of constant r have smaller circumferences than flat space would imply, $2\pi R_c \sin(r/R_c)$ instead of $2\pi r$.

The opposite case, one in which circles are larger than expected, is also interesting. We write

$$d\ell^2 = dr^2 + R_c^2 \sinh^2(r/R_c)\ d\phi^2$$

Such a metric is known as a hyperbolic geometry. Here, we are using the hyperbolic functions:

$$\sinh x = \frac{e^x - e^{-x}}{2}$$

instead of

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i}$$

.

see figure on slides.

This is similar to the sphere in that one can translate within the manifold and recover the same metric. Unfortunately, this manifold does not have a simple shape in 3-dimensions. The extra circumference makes one imagine a plane crinkled like a saddle, however this is only a local approximation. The key point is that the directions of curvature are in opposite directions. Unlike the sphere, r extends to infinity and the space have infinite area. We will concentrate on the spherical case, the hyperbolic case is completely symmetric to the spherical case, just change sin to sinh (sinch, or shine).

Now, let's consider 3-dimensional spaces. The flat space metric is easy

$$d\ell^2 = dx^2 + dy^2 + dz^2$$

In spherical coordinates, we have

$$d\ell^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\ d\phi^2)$$

Again, the 3-sphere of radius R_c in 4 dimensions has full homogeneity. The metric is

$$d\ell^2 = R_c^2(d\chi^2 + \sin^2\chi\ d\theta^2 + \sin^2\chi \sin^2\theta\ d\phi^2)$$

If we choose $r = R_c\chi$, we reach

$$d\ell^2 = dr^2 + R_c^2 \sin^2(r/R_c)\left[d\theta^2 + \sin^2\theta\ d\phi^2\right]$$

The term in brackets looks like the standard angular distance in spherical coords. Again, as before, if $r \ll R_c$, we recover the flat space metric, whereas at larger r the circumference is smaller than one expects. We will concentrate on the spherical case, the hyperbolic case is completely symmetric to the spherical case, just change sin to sinh.

These are the only homogeneous 3-dimensional spaces! Of course, one can always change coordinates to disguise the familiar form (one the homework problems), but any homogeneous space can be brought to one of these three forms.

We have now generated some coordinate systems for homogeneous spaces. Because we want to discuss time evolution, we must get time into the picture!

In special relativity, we made use of a new length in space-time.

$$ds^2 = dt^2 - \frac{d\ell^2}{c^2}$$

We thought of the motions of particles as curves in space-time in which the time direction was treated specially in the metric. Recall that the distance ds is invariant under Lorentz transformations.

Photons move as $ds = 0$. Massive particles have $ds > 0$.

We want to include the expansion of the universe. For this, we will choose coordinates that move with the particles. In other words, a given particle that is at rest with respect to the expansion will have constant coordinates. And the expansion of the universe is reflected in the change of the size/scale of the universe as a whole. This coordinate system is called a comoving system, In this system, each galaxy is labelled by r, which is called the comoving radial distance coordinate. It is fixed, i.e., not changing with time. Note that we are dealing with isotropic and homogeneous universe here, so there is no peculiar velocity. The expansion of the universe is defined in a quantity $R(t)$ or a(t), the scale factor.

$$dl = R(t)dr$$

In these coordinates, the spatial metric is

$$d\ell^2 = R(t)^2 \left[dr^2 + R_c^2 \sin^2(r/R_c) \left(d\theta^2 + \sin^2\theta \, d\phi^2\right)\right]$$

where $R(t)$ will simply change all the distances by some function of time. Note that R_c is time-independent (Longair calls this $\mathcal{R}$). It is conventional to pick $R(t) = 1$ at the present day, so the coordinates reflect the present-day scale of the universe.

This is called the comoving frame.

The full space-time metric is then

$$ds^2 = dt^2 - \frac{R(t)^2}{c^2} \left[dr^2 + R_c^2 \sin^2(r/R_c) \left(d\theta^2 + \sin^2\theta \, d\phi^2\right)\right]$$

This is known as the Robertson-Walker metric. Remember that sin goes to sinh in the hyperbolic case. And note that when we ask R_c, the radius of the curvature goes to infinity, then everything goes back to the flat case.

We will concentrate on the spherical case, the hyperbolic case is completely symmetric to the spherical case, just change sin to sinh.

It is also common for $R(t)$ to be written $a(t)$. It is called the expansion factor.

This is a very important formula! From this, one calculates all distances and volumes in cosmology. Note that while can compute between any two points, it is often simplest to put one of the points at the origin $r = 0$. This avoids the problem of solving for the equation of straight line.

We have not yet specified the expansion factor $R(t)$. Computing it requires specifying all the gravitating mass in the universe. For now, we leave it general because one can compute a number of important quantities for arbitrary $R(t)$.

Lecture 4. Measurements

Next, let's consider measurements in cosmology, we will go over: **redshift, distance, volume, look-up time and horizon**. These derivations are not as complicated as their reputation would lead one to believe! You need to study the derivations to learn how to reproduce them.

3 Redshift

We will explain what is the cosmological redshift. Of course, this is related to what we all know as the Hubble's law. $v = cz = H_0 r$. What do we think Hubble's law is telling us? Or proving? GR? Big Bang? recall that Einstein realized that his steady state cosmology was incorrect after Hubble discovery, called it the biggest blunder of his career.

First, consider the propagation of photons. For photons, $ds = 0$. We now want to know what happens when we observe light from a distant galaxy at radial coordinate r. We will place ourselves at the origin $r = 0$ and consider the arrival of photons from a source at location r_1. Obviously, the photon travels purely radially in our coordinate system, so $d\theta = d\phi = 0$. We have

$$\frac{c}{R(t)}dt = -dr$$

where the minus sign is because the photon is traveling toward the origin. **this one remains on the board!** Here

$$r = \int C/R(t)dt$$

is the **comoving distance.**. Note again this is the coordinate, but can't actually measure.

Now let's think about the case we observe the light, as an EM wave, with a certain frequency. Let's imagine two signals sent to us. One leaves at t_1 and is received at t_0; the next is sent at $t_1 + \Delta t_1$ and is received at $t_0 + \Delta t_0$.

Let's integrate along the path to find the travel time.

$$\int_{t_1}^{t_0} \frac{c}{R(t)} dt = -\int_r^0 dr = r$$

The second signal is

$$\int_{t_1+\Delta t_1}^{t_0+\Delta t_0} \frac{c}{R(t)} dt = -\int_r^0 dr = r$$

Note that the coordinate distance traveled is the same; we have put the expansion of the universe in the $R(t)$ term, while the source and observer sit at fixed r. The real difference is that the universe as expanded a little since the first photon left.

Setting these equal gives

$$\int_{t_1}^{t_0} \frac{c}{R(t)} dt = \int_{t_1+\Delta t_1}^{t_0+\Delta t_0} \frac{c}{R(t)} dt = \int_{t_1}^{t_0} \frac{c}{R(t)} dt + \frac{c\Delta t_0}{R(t_0)} - \frac{c\Delta t_1}{R(t_1)}$$

So

$$\Delta t_0 = \Delta t_1 \frac{R(t_0)}{R(t_1)}$$

The difference of the arrival times has been dilated relative to the difference of the departure times by a factor that is the ratio of the expansion factors at the two times. You can understand it as time dilation in special relativity, but the way we derive it, this has nothing to do with relativity. If we imagine that our departure times were successive crests of an electromagnetic wave, then the period of the light must be altered by a factor $R(t_0)/R(t_1)$.

The light appears Doppler shifted. Its wavelength is changed by a factor $R(t_0)/R(t_1)$. We normally define the redshift by $\lambda_{obs} = \lambda_{emit}(1+z)$. This means that

$$1 + z = R(t_0)/R(t_1)$$

Again, the expansion of the universe causes this redshift without reference to gravity. One can think of the wavelength of the light being stretched by the expansion. Light from redshift $z = 1$ was emitted when the universe was half its present size.

If we consider the redshift of light reaching us, then our convention $R(t_0) = 1$ means that $1 + z = 1/R(t_1)$. This means that redshift can be thought of as a time variable $z(t)$.

Consider a galaxy close enough to us, that we don't need to worry about the relativistic term, or curvature yet. Locally, we like to think of the redshift as a velocity effect. For example, the Hubble law is written as a ratio of velocity to distance.

$$\int_{t_1}^{t_0} \frac{c}{R(t)} dt = -\int_r^0 dr = r$$

If $t_0 - t_1$ is small compared to the time scales over $R(t)$ changes, then we have $c(t_0 - t_1) = R(t_0)r$, or $c\Delta t = r$, which makes sense!

Now for the velocities.

$$z = \frac{R(t_0)}{R(t_1)} - 1 \approx \frac{R(t_0)}{R(t_0) + (t_1 - t_0)\dot{R}(t_0)} - 1 = \frac{1}{1 - (t_0 - t_1)\dot{R}(t_0)/R(t_0)} - 1 \approx (t_0 - t_1)\frac{\dot{R}(t_0)}{R(t_0)}$$

For small redshifts, the velocity is $v = cz$. We have

$$v = r\frac{\dot{R}(t_0)}{R(t_0)} = rH_0$$

so the ratio of the velocity to the distance $R(t_0)r$ is $H_0 = \dot{R}(t_0)/R(t_0)$. It is very important to remember this definition of Hubble constant, the **expansion rate**, or first derivative of the expansion. One can define the Hubble constant as a function of time $H(t) = \dot{R}/R$. This relates the observed expansion to the behavior of $R(t)$. So now we see that Hubble's law doesn't directly require big bang, or even GR, or any gravity law. It is a result of the metric we used, or a result of C.P. In fact, similar solutions exist for such explosion blast wave calculation. It is interesting, in historical context, how the observation of Hubble's law so directly supported modern cosmology.

4 Distance

What distances do we actually measure in cosmology? We have seen one, the **comoving distance** traveled in a given time

$$\int_{t_1}^{t_0} \frac{c}{R(t)} dt = -\int_r^0 dr = r$$

Note that if we change integration variables from t to $z = 1/R(t) - 1$, then we have $dz = -(dR/dt)/R^2\, dt = -H(t)(1 + z)\, dt$. So

$$r = \int_0^z \frac{c\, dz}{H(z)}$$

We will eventually have simple formulae for $H(z)$, but it's arbitrary for now. **remain on the board!**

However, we will find that this distance isn't necessarily the one that we measure! In fact, we can't measure r, the comoving distance at all.

How do we measure distance in cosmology? We can't go there and have a ruler. **this is an important point: classical cosmology is almost all about measuring distances.**

In astrophysics, we more or less use a combination of two ways. First, if we know the true angular diameter of something by other means, and we can measure the angular diameter observed using telescope, then we get the distance. Examples are like the definition of parallax, and the moving cluster method. This is called **the angular diameter distance**, $D_A = d/\Delta\theta$.

Second, if we know the true luminosity of something by other means, and we can measure the apparent luminosity, then we get the distance. Examples are like Cepheids, T-F etc. This is called **the Luminosity Distance**. $f = L/(4\pi D_L^2)$.

In other classes, you have dealt with measuring distances of stars, star clusters and galaxies. We will visit those briefly when we discuss Hubble constant next week.

Now let's see how to measure these distances in a cosmological context.

Consider an object of some known physical diameter at some redshift. What angle on the sky do we measure?

We will imagine ourselves at $r = 0$ and consider the object to be extended in the θ direction. In this case, our object is lying tangentially, so $dr = d\phi = 0$. Its diameter is then

$$d = R(t)R_c \sin(r/R_c)\Delta\theta$$

We would expect that $d/\Delta\theta$ would be the distance to the source. We define this ratio as the "angular diameter distance" D_A.

$$D_A = R(t)R_c \sin(r/R_c) = \frac{R_c \sin(r/R_c)}{1+z}$$

We computed $r(z)$ above.

In the flat cosmology, $D_A = r/(1+z)$. Note that the expansion causes the transverse distances to differ from the radial ones! The distance is smaller, or the object looks bigger than what the coordinate distance implies. Again, nothing to do with gravity. This is a result of expansion.

Next, let's consider a source of some luminosity L at some redshift z. We would expect the flux to be $L/4\pi D^2$. What do we actually get?

Place the source at the origin. How much solid angle does our telescope subtend? Now the question reversed, it is the angle that the distant galaxy will measure now. The metric says the transverse distance today corresponding to a given angle is $R(t_0)R_c \sin(r/R_c)\Delta\theta$. So the solid angle of an area A is $A[R_c \sin(r/R_c)]^{-2}$. The flux we receive per unit area is then $f = L/(4\pi[R_c \sin(r/R_c)]^2)$.

Is that all? There are other effects, however. The source is emitting a given number of photons per second, but we receive these photons in a time that is stretched by $R(t_0)/R(t_1)$. So the number flux is reduced by $1+z$. Moreover, each photon is redshifted and has its frequency reduced by $1+z$. So the energy flux is reduced by $(1+z)^2$.

We receive a flux

$$f = \frac{L}{4\pi[R_c \sin(r/R_c)]^2(1+z)^2}$$

If we define the "luminosity distance" by $f = L/4\pi D_L^2$, then

$$D_L = R_c \sin(r/R_c)(1+z) = D_A(1+z)^2$$

Warning: if one is concerned with the flux per unit frequency, then $f_\nu \neq L_{\nu(1+z)}/4\pi D_L^2$. The flux over some range in frequency does scale as D_L^{-2}, but the range of frequency itself is scaling as $(1+z)^{-1}$. So $f_\nu = L_{\nu(1+z)}(1+z)/4\pi D_L^2$.

Luminosity distance is larger than the coordinate distance, and larger yet than the angular diameter distance, as a result of expansion. In order words, object becomes faint very fast, but not much smaller. This has important consequence!

If the flux received goes as $f \sim 1/D_L^2$ while the angle subtended goes as $\Omega \sim 1/D_A^2$, then the surface brightness is going as $SB \sim f/\Omega \sim D_A^2/D_L^2$. This always goes as $(1+z)^{-4}$ without any reference to the density of the universe. High-redshift objects have lower surface brightness. This has important consequence in high-redshift galaxy observations!

There are a couple of other distances that can be measured, but they don't occur very often. The luminosity distance and angular diameter distance are the total major ways that we express distance in cosmology. Note that because the expansion of the universe, and the fact that how we define distance depends on the expansion in different ways, there are many different distances in cosmology, and for anything with substantial redshift, say $z > 0.1$, they are different. So when you say something is N Mpc away, it is not meaningful unless you see what it means. However, the true importance of talking about these distances is that given redshift, it gives you the relation between size and angular diameter, between apparent and absolute flux. That's what you really need when you are measuring a distant object and try to interpret the data!

5 Volume

How much comoving volume is there in a given solid angle and redshift range? Here we are asking comoving volume, so we can forget about the scale factor $R(t)$ for the moment.

$$dV_{com} = R_c^2 \sin^2(r/R_c) d\Omega dr$$

from the metric. Why we need the volume? Because we want to get the spatial density of, say, galaxies. So imagine that we carry out a survey, and we count the galaxies within a redshift interval dz, and a solid angle $d\Omega$, we want to know how to get the density of these galaxies, so we need to understand how $dz \times d\Omega$, which is the thing that we measure. It is related to dV, which is the thing that we need for calculating volume.

Differentiating

$$r(z) = \int_0^z \frac{c\ dz}{H(t)}$$

gives $dr/dz = c/H(t)$.

$$\frac{dV_{com}}{d\Omega dz} = \frac{R_c^2 \sin^2(r/R_c)c}{H(t)}$$

6 Look-Back Time

Finally, we would like to calculate the look-back time to a given redshift, the time from a certain redshift to the current epoch, which is something we are clearly interested in, for example, to calcuate how much time there is for the stars in a galaxy to evolve if it was formed at $z = 5$.

From $dt = -dr\ R(t)/c$, we have

$$t(z) = \int_r^0 \frac{dr}{c} R(t) = \int_0^z \frac{dz}{H(t)} \frac{1}{1+z}$$

7 Horizons

Horizons. Here the question whether this is a maximum distance that we could see, or could influence us, from the past, and where light from us will reach in the future. The concept of **particle and event horizons.**

Returning to the propagation of light in our metric, it is possible that the integral

$$r = \int_0^z \frac{c\ dz}{H(t)}$$

converges to a finite answer as $z \to \infty$. This is the beginning. This means that only a finite volume of the universe is within our past light cone. In other words, there are spatial points from which we cannot have received light and hence cannot be affected by causal physics. This effect is called a **particle horizon**, it is asking the question: since the universe has a beginning, and it takes time for light to travel, which is the fastest way to establish causal communications, what is the maximum distance over which causal communication could have taken place at a certain epoch. In other words, what is the distance a light signal could have travelled from the origin of the big bang (if there were one) at t=0, by the epoch t. In a few lectures, when we are discussing the generation of CMB, we will encounter a rather severe problem for cosmologist, namely, the particle horizon size at the time when CMB was generated (z=1100) is only 2 deg or so on the sky, on the other hand, CMB is very uniform across the sky. So the question, which as some of you might know, gave rises to the idea of inflation, is, then, how one side of the universe were able to coordinate with the other side of the universe and had the same T, if there were no chance for any causal communication.

There is another possible horizon in which we compute the radius a light ray can reach in the infinite future. If

$$\int_{t_0}^{\infty} \frac{c\;dt}{R(t)}$$

is finite, then news of events beyond that radius will never reach us and vis versa. This is called an **event horizon.** In black holes, the event horizon is the surface into which we cannot gain information from inside. In this case, the event horizon is an enormous sphere, and we cannot get information from outside of it! In particular, in a Λ-dominated universe, as we will find out, the universe accelerate exponentially in later times, and this integral can be finite, which means that we will eventually be in some kind of isolated bubbles. It also has important consequence in cosmology.

We have now studied the Robertson-Walker metric for a general $R(t)$. In fact, the behavior of $R(t)$ depends on the gravitational attractions of the homogeneous matter. In the next two lectures, we will introduce the basic equation for the dynamical expansion of the universe, the Friedmann equation. It is the expansion that obeys GR.

8 Newtonian Analogue

Usually, one derives this from General Relativity, but let us first consider a Newtonian toy problem. Consider a homogeneous density distribution in uniform expansion. If we pick a origin and draw a sphere around it, then we might calculate the gravitational force by

Gauss's law. This means that the material outside the sphere doesn't contribute. In detail, we haven't proven this (nor can we).

We will let the radius of the sphere move with its boundary particles. The radius of the sphere is denoted $R(t)$. The mass inside it is constant. If the density today is ρ_0 and the radius today is R_0, then the mass is $4\pi\rho_0 R_0^3/3$. The acceleration is then

$$\ddot{R} = -\frac{4\pi}{3} G\rho_0 \frac{R_0^3}{R^2}$$

As the sphere expands, the density must scale as $\rho = \rho_0 R_0^3 R^{-3}$. So we also have

$$\ddot{R} = -\frac{M}{R^2} = -\frac{4\pi}{3} G\rho R$$

This is a second-order ODE for the expansion of our toy universe. We must specify two boundary conditions, which we will take as the radius and Hubble constant today.

$$R|_{t_0} = R_0$$

$$(\dot{R}/R)|_{t_0} = H_0$$

Using a familiar DE trick:

$$\ddot{R} = d\dot{R}/dt = d\dot{R}/dR\dot{R}$$

We can integrate this equation and get:

$$\frac{1}{2}\dot{R}^2 = \frac{4\pi}{3} G\rho_0 \frac{R_0^3}{R} + C/2$$

Here C is a constant of integral. This ODE is the famous Friedmann equation, which provides the evolution of the scale factor. It is the dynamical equation of the evolution of the universe. Think of it as the Newton's law in cosmology. In fact, we just derived from Newton's law in a homogeneous sphere case. It can be derived from GR, Einstein's field equation, assuming cosmological principles, and the RW metric. Note that this equation has the form of an energy equation, as if C is some sort of total energy. And as we shall see, C determines the fate of the universe, whether it has enough energy to expand forever, or it is bound and will collapse again. We like to write this energy equation in the form of Friedmann equation:

$$\dot{R}^2 = \frac{8\pi}{3} G\rho R^2 + C$$

Remember that the Hubble constant is $\dot{R}/R$ evaluated today. This means that today

$$H_0^2 = \frac{8\pi}{3} G\rho_0 + \frac{C}{R_0^2}$$

This sets the constant C.

Let us define a constant $\Omega_0 = 8\pi G\rho_0/3H_0^2$. Then we have

$$1 = \Omega_0 + \frac{C}{R_0^2 H_0^2}$$

$$C = R_0^2 H_0^2 (1 - \Omega_0)$$

Let's put this back in the equation for $\dot{R}$. We can write it as

$$\dot{R}^2 = \frac{8\pi}{3} G\rho_0 \frac{R_0^3}{R} + R_0^2 H_0^2 (1 - \Omega_0)$$

$$\left(\frac{\dot{R}}{R}\right)^2 = \Omega_0 H_0^2 \left(\frac{R_0}{R}\right)^3 + (1 - \Omega_0)\left(\frac{R_0}{R}\right)^2$$

$$\frac{H(z)}{H_0}^2 = \Omega_0 (1+z)^3 + (1 - \Omega_0)(1+z)^2$$

This is a simple formula for $H(z)$.

Note that in this equation, we are assuming that $\rho \sim (1+z)^{-3}$. This is true for matter, because the matter should be conserved. Relativistic matter will have to be treated differently.

Recall that we were able to express our distance-redshift relations in terms of integrals over $H(z)$. For example, we had the coordinate distance

$$r = \int_0^z \frac{c\, dz}{H(z)} = \frac{c}{H_0} \int_0^z \frac{dz}{[\Omega_0 (1+z)^3 + (1 - \Omega_0)(1+z)^2]^{1/2}}$$

9 Friedmann Equation

However, our toy model is missing an important aspect, which is the dynamical effect of the curvature of the universe. For this, we need to appeal to GR. GR expresses gravitational forces as curvatures in space-time, so that particles that are not being acted on by non-gravitational forces fall along straight lines in space-time that appear as accelerating trajectories. Recall that the curvature of a manifold is completely specified by the (spatially-dependent) metric.

We will not prove it, but GR yields the following two equations for the evolution of $R(t)$.

$$\ddot{R} = -\frac{4\pi G}{3} R \left(\rho + \frac{3p}{c^2}\right) + \frac{1}{3}\Lambda R \quad \text{(Note typo in Longair!)}$$

$$\dot{R}^2 = \frac{8\pi G}{3} \rho R^2 - \kappa c^2 + \frac{1}{3}\Lambda R^2$$

$$\kappa = 1/(R_c^2)$$

These look similar to the above, but with some new terms. First, the constant of integration in the Friedmann equation has been replaced by a term depending on the curvature κ. Second, we have terms that depend on pressure: matter: $p = 0$, and photon: $p = \rho c^2/3$, and on the cosmological constant Λ.

This Λ is something that you might remember if you still remember your Einstein's field equation

$$R_{\mu\nu} - 1/2 g_{\mu\nu} R - \Lambda g_{\mu\nu} = 8\pi G/c^4 T_{\mu\nu}$$

where $R_{\mu\nu}$ is the Ricci tensor, and R is the curvature scalar. This is the equation, of course, ties the geometry of the universe $g_{\mu\nu}$ with the stress-energy tensor $T_{\mu\nu}$ that describes density and energy/momentum,. And recall that Einstein's the original purpose was to introduce a term to stabilize the field equation to avoid the expansion solution.

Note what this Λ will do to your equation: (1) it is a replusive force in the acceleration equation, while matter and pressure is always attracting, gravity slows things down. In other words, you need Λ to balance things. (2) by carefully choosing Λ and κ, you can have a stable solution, where $\ddot{R} = \dot{R} = 0$, at this stable radius $R = R_s$. This is called the Einstein-Lemaitre model, this is the reason why Einstein chose to use Λ in the original field equation, to obtain a stable solution, because it is very clear from these equations that otherwise the universe would have to expanse or contract, and have to decelerate.

Let's ignore p and Λ for now. The second equation gives us

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} - \kappa c^2/R^2 = \frac{8\pi G\rho_0}{3}(1+z)^3 - \kappa c^2(1+z)^2$$

We define $\Omega_0 = 8\pi G\rho_0/3H_0^2$ and find

$$\left(\frac{H(z)}{H_0}\right)^2 = \Omega_0(1+z)^3 - \frac{\kappa c^2}{H_0^2}(1+z)^2$$

At the present-day ($z = 0$), the LHS is 1, so we must have

$$\Omega_0 - 1 = \frac{\kappa c^2}{H_0^2}$$

This says that the value of Ω_0 is directly related to the curvature of the universe. One does not have the freedom to pick an arbitrary geometry, density, *and* Hubble constant!

If $\Omega_0 = 1$, then $\kappa = 0$ and the geometry of the universe is flat. If $\Omega_0 < 1$, then $\kappa < 0$ and the geometry is hyperbolic (open). If $\Omega_0 > 1$, then $\kappa > 0$ and the geometry is spherical (closed).

10 Solution

Before we solve the Friedmann equation, let's look at its behavior. We can rewrite the Friedmann equation as:

$$\dot{R}^2 = H_0^2[\Omega_0(1/R - 1) + 1]$$

when $R \to \infty$,

$$\dot{R}^2 = H_0^2(1 - \Omega_0).$$

For $\Omega_0 = 1$ case, at $R = \infty$, $\dot{R} = 0$, so the universe will expand forever, but it eventually will come to a stop. This case is called the Einstein-de Sitter model, or the critical model.

For $\Omega_0 > 1$ or $\kappa > 0$ case, at some large R, $\dot{R}$ will go zero, so the universe will reach its max radius, and then begin to collapse. The expansion will never reach infinity and it will ends in the big crunch, this is the close model.

For $\Omega_0 < 1$ or $\kappa < 0$ case, $\dot{R}^2$ is always positive, the universe will expand forever, this is the open model.

Hence, in this simple formulation, we have direct connections between the density of the universe (as measured by Ω_0), the geometry of the universe, and the fate of the universe! However, these relations do breakdown when we introduce Λ. $\Omega_0 = 1$ is special. The required density $\rho_c = 3H_0^2/8\pi G$ is known as the critical density. $\Omega_0 = \rho_0/\rho_c$.

The ODE of Friedmann equation, when $\Lambda = 0$, can be solved analytically. $\Omega = 1$ case is easy, in this case:

$$\dot{R}^2 = H_0^2/R,$$

you can show easily that:

$$R = (t/t_0)^{2/3}$$

so the universe grow as a power law. And comparing these two equations, you can show easily:

$$t_0 = (2/3)H_0^{-1}.$$

Therefore, it reveals another physical meaning of Hubble constant: the inverse of H_0 gives the age of the universe.

Obviously, for $\Omega >$ and < 0, the universe is closed, or open, and grows slower and faster than the critical case, respectively. For open model, there is a simple case where the universe is completely empty, this is called the Milne model: $\Omega_0 = 0$, in this case, there is obviously no deceleration, and the universe will expand with a constant speed, $R(t) = H_0 t$.

Otherwise, the solution of Friedmann equation can be most conveniently written in parametric form: For $\Omega_0 > 1$:

$$R = a(1 - \cos\theta), t = b(\theta - \sin\theta),$$

$$a = \frac{\Omega_0}{2(\Omega_0 - 1)}, b = \frac{\Omega_0}{2H_0(\Omega_0 - 1)^{3/2}}.$$

Obviously, R will reach its max when $\theta = \pi$, and $t_{max} = \frac{\pi\Omega_0}{2H_0(\Omega_0-1)^{3/2}}$.

Similar solution can be found for the open model.

See figure on slides.

The relevance of the closed model solution is: as we will see later when we discuss galaxy formation, for a local area in the universe where the density is higher than the critical density, it can be treated like a closed universe, and it will collapse at this t_{max} timescale.

We can rewrite Friedmann equation:

$$\dot{R}^2 = \frac{8\pi G}{3}\rho R^2 - \kappa c^2 + \frac{1}{3}\Lambda R^2$$

by dividing $\dot{R}^2$ in both sides, and ask: $\Omega_Lambda = \Lambda/3H_0^2$, and $\Omega_k = -kc^2/H_0^2$, thus:

$$\left[\frac{H(z)}{H_0}\right]^2 = \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda.$$

This is the form of Friedmann equation I find most useful, because it connects Hubble constant with other cosmological parameters: density, curvature and cosmological constant. Obviously, at $z = 0$:

$$1 = \Omega_m + \Omega_k + \Omega_\Lambda,$$

a relation that has been referred to as cosmic triangle. We will discuss how to fill in this cosmic triangle in the next week.

11 Λ model

The solution for Λ case is more complicated. But let's look at Friemann equation again, if $\Lambda \neq 0$,

$$\dot{R}^2 = \frac{8\pi G}{3}\rho R^2 - \kappa c^2 + \frac{1}{3}\Lambda R^2$$

There are three terms: the gravity term, curvature term and Λ term. See how different terms change with redshift. At $R >> 1$, the curvature term doesn't grow with R, and can be ignored; the gravity term goes slower than the Λ term, and can be ignored as well. So we will have that in the late stage of expansion in a Λ universe,

$$\dot{R} = \sqrt{\Lambda/3}R, R \propto \exp(\sqrt{\Lambda/3}t)$$

The universe is going to expand exponentially if $\Lambda > 0$! Note that this is true regardless of the curvature and as far as $\Lambda > 0$, even if it is relatively small, you can not escape the fate that the universe is going to expand exp, and it seems that we can't avoid this fate now. So very small (many billion years later), the universe is going to be exceedingly empty. We will meet this exp expansion again twice in this class, one about the dark energy, the other about inflation.

12 flatness problem

So how about when R is very small, in other words, at high redshift? For matter, we have $\rho \propto R^{-3} \propto (1+z)^3$, this term is going to dominate the Friedmann equation at early epoch, since the gravity term will go as $(1+z)$, curvature term is a constant, and Λ term goes as $R^2 \propto (1+z)^{-2}$. If we ignore the last two terms, then

$$\dot{R}^2 = \frac{8\pi G}{3}\rho R^2,$$

and divide this by R^2, and note that $\rho_c = 3H^2/8\pi G$, we have $\Omega = \rho/\rho_c = 1$ at high redshift, regardless of what the current day curvature and cosmological constant is. There are two things to learn from this:

1. at high-redshift, the universe can always be approximated as being an Einstein-de Sitter model. Given the range of cosmological parameters we know today, the universe can be regarded as more or less being flat at $z > 4$ or so.

2. what is called the flatness problem. That is: if the curvature of the universe we measure today is close to zero, or Ω_0 is within a factor of 10 from unity, then at the very beginning, the univese must have been fine tuned to be very close to completely flat. Because if there is any very small deviation from flat at high redshift, it will be blow up more or less by a factor of $1+z$ now. The flatness of the universe is a strong function of z, so why we happen to be living at a time Ω_0 is not far from 1, while if the curvature is not zero, then Ω_0 can be equally possible to be 1000 or 1/1000? Come back to cosmological principle, this argues that if we believe that we are not living in a special time, then the universe must have always been precisely flat from the beginning to the end. This argument, or solution, for the flatness problem, is a strong motivation for inflation model, in fact, you can say that

it is a prediction for such inflation model, because at that time, there was few evidence that the universe is flat. We will discuss this again later.

Now let's consider the relation between the $\ddot{R}$ and Friedmann equations. Differentiating the Friedmann equation gives

$$\ddot{R} = \frac{8\pi G}{3}\rho R + \frac{4\pi G}{3}R^2\frac{\dot{\rho}}{\dot{R}} + \frac{1}{3}\Lambda R$$

Subtracting the original GR acceleration gives

$$R\frac{\dot{\rho}}{\dot{R}} = -3\left(\rho + \frac{p}{c^2}\right)$$

We can write this as

$$R\frac{d\rho}{dR} = -3\left(\rho + \frac{p}{c^2}\right)$$

where we consider ρ as a function of R.

We are familiar with normal matter. Here, ρ is the familiar density and $p \ll \rho c^2$. So $\rho \propto R^{-3}$ as we have used.

What about other kinds of matter? Let's write the equation of state

$$p = p(\rho).$$

Photons have a pressure that is a third of their energy density. We write $p = \rho c^2/3$. That means that $\rho \propto R^{-4}$. This corresponds to the number density of photons dropping as R^{-3} while their frequencies and hence energies decrease as R^{-1}.

In the case when $\rho \propto (1+z)^4$, it is easy to show that

$$\dot{R}^2 = 8\pi G/3\rho R^2, \dot{R}^2 \propto 1/R^2$$

or

$$R \propto t^{1/2}$$

We often separate ρ into two pieces, the non-relativistic matter and the relativistic photons, neutrinos etc. We define $\Omega_M = \rho_m/\rho_c$ and $\Omega_r = \rho_r/\rho_c$ so that

$$\frac{\rho}{\rho_c} = \Omega_m(1+z)^3 + \Omega_r(1+z)^4$$

Note that since radiation goes as $(1+z)^4$, at some point it will dominate, so at very high redshift, it is always radiation dominated, and you can ignore mass when solving the dynamical evolution of the universe, while at later time, it is always matter dominated, as it is now. But in CMB era, it is radiation dominated. It is very important to remember that for matter: $\rho \sim (1+z)^{-3}$, and for radiation: $\rho \sim (1+z)^{-4}$.

13 Deceleration Parameter

Let's introduce one more parameter. We showed the Hubble constant is the expansion rate of the universe. So how do we define the accerelation, or more precisely, the deceleration, of the universe. We define:

$$q_0 = -\left(\frac{R\ddot{R}}{\dot{R}^2}\right)_{t_0}.$$

Subsituting this definition to basic equations, when without Λ, and without pressure term, it is immediately obvious that

$$q_0 = \Omega_0/2.$$

So this shows that the deceleration of expansion is directly related to the mass density. Not to be surprised, the deceleration is caused by the gravity. But what it gives us is another way to measure the density of the universe, i.e., we know the expansion rate of the universe today; if we can somehow measure the expansion rate of the universe at high-redshift, then we can measure the deceleration and thus the density. This flavor of measuring Ω is called the geometrical measurement.

Note that in the absense of Λ, the universe always decelarate. But the above relation is not valid in case of Λ. As you probably know, the universe is actually accrelate.

14 Observations

A few lectures ago, we discussed the observables in cosmology, such as the different kind of distance, the comoving, angular diameter, and luminosity distances, the volume, and look-back time. We derive their basic relation with RW metric, but since at that time we didn't know R(t), we can't write down their relation with redshift. Now we can.

If we set $\Lambda = 0$, then we have

$$H(z) = H_0\sqrt{\Omega_0(1+z)^3 - \kappa c^2(1+z)^2} = H_0\sqrt{\Omega_0(1+z)^3 + (1-\Omega_0)(1+z)^2}$$

Let's use this to find our distance-redshift relations.

$$r = \int_0^z \frac{c\,dz}{H(z)} = \frac{c}{H_0}\int_0^z \frac{dz}{\sqrt{\Omega_0(1+z)^3 + (1-\Omega_0)(1+z)^2}}$$

The integral is a bit of a mess and the expression depends on whether $1-\Omega_0$ is positive or negative.

The next step is to calculate $D = R_c \sin(r/R_c)$ (or sinh), because we found that $D_A = D/(1+z)$ and $D_L = D(1+z)$. This yields

$$D = \frac{2c}{H_0 \Omega_0^2 (1+z)} \left\{ \Omega_0 z + (\Omega_0 - 2) \left[(\Omega_0 z + 1)^{1/2} - 1 \right] \right\}$$

for any Ω_0.

Hence, we have just found the relation between distances and redshift. For small z, we have $D = (c/H_0) \left[z - z^2(1+\Omega/2)/2 + \ldots \right]$

c/H_0 appears a lot in cosmology. It is known as the Hubble distance. It is $3h^{-1}$ Gpc. Astronomers like to write $H_0 = 100h$ km/s/Mpc.

What do these distance relations look like? Let's consider $\Omega = 1$. Here

$$D = (2c/H_0) \left[1 - \frac{1}{\sqrt{1+z}} \right]$$

As $z \to \infty$, $D \to 2c/H_0$.

However, the angular diameter distance is $D/(1+z)$ so this actually goes to 0! The maximum D_A is at $z = 5/4$ and is $8c/27H_0$. The angular diameter distance for $0.5 < z < 5$ is always about $1h^{-1}$ Gpc. This means that $1''$ is about $5h^{-1}$ kpc (physical) at cosmological distances.

You can do the same for open cosmology or cosmology with a Λ. You need to use a lot of these calculations when working on cosmology, and when working on extragalactic astronomy in general. I recommend a few tools: (1) David Hogg's paper (astro-ph/9905116); (2) astropy.cosmology; (3) Ned Wright's cosmology calculator, http://www.astro.ucla.edu/ *wright*/CosmoCalc.h (4) iCosmos, http://www.icosmos.co.uk/; (5) Cosmology calculation i-phone app CosmoCalc, free. Homework 2, which will be assigned next week, is to ask you to write your own astropy.cosmology.

The luminosity distance, however, grows quickly with z. High-redshift objects may not get smaller, but they do get much fainter!

What are these relations used for? (1) given a cosmology, and measured redshift, you can then derive the diameter, luminosity and time of a high-redshift source from the observed angular size, flux, and redshift.

(2) if you can get the diameter, luminosity etc. independently, then you can use these figures to constrain cosmology, like Hubble constant, density, cosmological constant etc.

Notes on Cosmological Parameters
Feb 2024

In matter dominated universe, we have:

$$\left[\frac{H(z)}{H_0}\right]^2 = \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda.$$

This is the form of Friedmann equation I find most useful, because it connects Hubble constant with other cosmological parameters: density, curvature and cosmological constant. Obviously, at $z = 0$:

$$1 = \Omega_m + \Omega_k + \Omega_\Lambda,$$

What Friedmann equation gives us is the expansion history of the universe. Through it, we also introduced a number of cosmological parameters.

- $H_0 = \dot{R}/R|_0$, is the current expansion rate of the universe.
- t_0 is the current age of the universe since the BB.
- Ω_m is the current density parameter of the Universe.
- k is the curvature of the universe, deciding the geometry.
- Λ is the cosmological constant.
- q_0 is the deceleration parameter.

Clearly, they are not all independent. Indeed, they are tied by Friedmann equations. In matter dominated era, as the equation we showed above, the entire expansion history can be described by three parameters, including a scale (Hubble constant, or age of the universe), and two parameters that specify the relative contribution of matter (including dark matter), curvature and cosmological constant to the total energy-density budget of the universe. This is our **Robertson-Walker-Friedmann world model.** The most important task of a cosmologist is to understand what our world model is. And our biggest task is to test whether this world model, and which version of it, is supported by our observations. This is called the **cosmological tests.**

In fact, our cosmological test, or our cosmological model, includes even more parameters, because we are interested in not only the expansion history of the material dominated era, but (1) the state of the universe in radiation dominated era, i.e., CMB, (2) the growth of fluctuation in the universe. We will visit those later in the semester.

Cosmological tests:

- **Expansion history**
- **Radiation-dominated era – anisotropy of CMB**
- **growth of fluctuation – galaxy clustering, IGM**

This figure shows the number of parameters presented by the Planck summary. The main conclusion of this is that our cosmology can be described by essentially 6/7 parameters that fit all existing observations satisfactorily. These observations are (almost all):

(1) the expansion history of the universe, including measurements such as supernova, which we will discuss next lecture;

(2) the anisotropy of the CMB;

(3) the large scale structure in low-z universe, including measurements of galaxy clustering and structure of the IGM, as well as clusters of galaxies

And all of them can be fit by seven parameters (paper said six, because it is assuming a flat geometry: $H_0 = 70$, $\Omega_m = 0.28$, $\Omega_b = 0.05$, $k = 0$, and then two parameters, the zero point and slope that specify the density fluctuation power spectrum, and one parameter describes the cross section of free electrons to CMB that will affect CMB photons through Compton scattering. We will discuss Ω_b, and the other three later. For the moment, we will concentrate on the three parameters that appear in Friedmann equations, i.e., the expansion history of the universe, which we generally call the **classical cosmological tests**. But the cosmological tests will really dominate our discussions in this class in the next 1.5 months of the class. We will discuss how to measure Ω_b in the BBN lecture, and discuss CMB in details, as well as large scale structure tests in details later.

1 Hubble Constant

We already heard from Matthew about history of Hubble Constant.

The distance scale path has been a long and tortuous one, but with the imminent launch of HST there seems good reason to believe that the end is finally in sight. – Marc Aaronson

The determination of the distance scale of the universe, the Hubble constant, has been one of the most fascinating chapter of modern astronomy. ALMOST all of last century's great astronomers had a role, both positive and negative, on it. It involves a large amount of astrophysics, and a large amount of resources, it is controversial and had many ups and downs. It had a lot of characters, a lot of stories that we can tell. Some of these are older stories now, but you will get a sense of how our field developed. For the stories, I will refer you to an old, excellent book, 'The Lonely Hearts of the Cosmos", by Dennis Overbye. The central scheme of the book is about the determination of Hubble constant.

Hubble constant has sometimes been called the most variable constant in science. As an example of that, look at (Huchra Figure 1). A bit of history, the first published determination of Hubble constant is not from Hubble himself, but from Lemaitre in 1927 based on Hubble's observations, and it is still the largest value to date, about 630 km/s/Mpc. Hubble himself finally weighted in in 1929 at 500. Also, very early on, the Dutch astronomer, Jan Oort, thought something was wrong with Hubble's scale and published a value of 290 km/s/Mpc, but this was largely forgotten.

The first major revision to Hubble's value was made in the 1950's due to the discovery of Population II stars by W. Baade. To distinguish pulsating variables RR Lyr (PopII) with Cepheids (PopI). That was followed by other corrections for confusion, etc. that pretty much dropped the accepted value down to around 100 km/s/Mpc by the early 1960's.

But after that, it entered an era of a very and intense debate. (Huchra Figure 2). So what is shown here is the Hubble constant determination from 1970 on. Note two symbols here, (1) open stars are values determined by two Gerard de Vaucouleurs and Sidney Van Ben Bergh. At least until 1990s, they always got H=100. (2) star symbols are values determined by another prominent group, by Allan Sandage, Hubble's student and hand-picked heir, who might be the most important astronomer after Hubble, and his friend Gustav Tammam, a Swiss astronomer. For about 40 years, their values were always 50 - 55.

Both can't be right.

Then there are the third group of people, the people who initially didn't not belong to these two camps. They are mostly "young" astronomers, at least when this group of astronomers entered the game in the 1970s. They invented a whole new arrays of distant indicators, and were much more adapted to the new technology at the time, such as CCDs. Their values more or less lied in the middle, like 75.

Then came HST. The single most important mission of HST is to measure Hubble constant, there is no doubt about that. The group that in charge of that is called the "Key Project" team. The measurement of Hubble constant is one of the two Key Projects of HST. Key project team used a combination of distance indicators, and did exclusively detailed work on the Cepheids in galaxies in Virgo clusters to calibrate their distance indicator. The final result of the HST Key Project, which was published about two years ago, was 72 ± 8. One of the most important goal of HST is to determine the Hubble constant within 10%, and they were able to do that.

I will argue that it is also an extraordinarily difficult constant to determine. Let's see why:

we have Hubble's law:

$$cz = v = Hr$$

so the recession velocity, or the redshift, is proportional to the distance of a distant object. In order to determine H_0, obviously, we need to determine two things: measure the redshift of a distant galaxy, and measure the distance to it with an independent method. So how far a galaxy can you use? We mentioned that the random motion of a galaxy due to its local gravitational field, i.e., the motion that has nothing to do with the Hubble flow, the uniform expansion of the universe, is of the order 500 - 1000 km/s. So in order to measure the Hubble constant that is not affected by this deviation from Hubble flow, we then have to go to sufficiently large distance when the velocity of the galaxy is dominated by the the expansion, not by the local velocity field. So in order to measure Hubble constant to an accuracy of 10%, we need to go to a expansion velocity of, say 7,000 km/s. If we assume that the Hubble constant is 70 km/s/Mpc, plug this back to Hubble's law, we find that it required us to measure the distance of a galaxy at 100 Mpc away. This is very far away.

The other difficulty we have is that although using relations such as Tully-Fisher relation, Faber-Jackson relation and likes, we can measure the distance of faraway galaxies, these distance measurements are all **Relative.** It means that these relations more or less tell us how to determine the proportionality of galaxy properties. For example, T-F tells us that $L \propto v^4$, so if you measure the rotation curve of two galaxies, you can then know the relative intrinsic brightness of these two galaxies. Therefore, if you know the distance of one galaxy, and the relative observed brightnesses, then you can get the distance of the other galaxy. But it still requires you to know at least one galaxy, and therefore the relation needs to be calibrated. There are two flavors of distance indicators. Most of the distance indicators that we know of that apply to distant galaxies are relative distance indicators, they need to be tied to an absolute distance indicator in order to be applied to get Hubble constant. This process is very non-trivial. The simplest absolute distance indicator we know is the familiar trigonometric parallax. But it can only be applied to objects within few tens of parsecs with sufficient accuracy. On the other hand, in order to calibrate such as T-F relation, we need to measure the absolute distance to at least a few nearby spiral galaxies. We only know two spiral galaxies, M31 and M33, hardly enough. So we need to go to the nearest clusters of galaxies, such as Virgo, and Fornax clusters, which are a few Mpc away. To tie the distance of galaxies a few Mpc away with stars at few tens of parsecs away is very difficult.

The classical procedure for estimating distances in the Universe typically uses a series of relative distance indicators; the distance of a nearby galaxy might be determined by comparing the apparent brightness of its individual stars to those of similar stars in the Milky Way; the distance of those Milky Way stars are compared to some more nearby stars that we can use absolute distance indicators, and so on. This bookstrapping approach to measuring distances is called the cosmic distance ladder, where each rung up the ladder takes us to greater distance, until we are out into the uniform Hubble flow to measure the distance indicated by the redshift.

You obviously DON'T want to have too many steps on the ladder. None of the distance indicator is perfect, they all have their errors, that is, the relation such as Cepheids, T-F, can not uniquely predict the distance, but only with a degree of uncertainty. When you begin to build the distance ladder, the error will build up, the relative error term more or less increases by the errors adding in quadrature,

$$(\sigma(D)/D)^2 = \Sigma(\sigma(D_i)/D_i)^2.$$

So if you require the final Hubble constant to be 10% accurate, and you have 10 steps to reach it from trigonometric parallax in your distance ladder, than you need each step to be about 3% accurate, which is very hard to achieve.

Because the error budget is so tight, there are two things that people should worry about, (1) to find absolute distance indicator (other than parallax) to minimize the number of steps; (2) to try to make sure that the result is not biased, no systematic errors.

The HST key project, whose goal is to measure H_0 with 10% accuracy, was formed in 1986, by Marc Aaronson, and started their observations in 1991. The final result of the key project is published in 2001. The key project way to determine H_0 is (1) to find and calibrate Cepheid distance in a large number of nearby (< 20 Mpc) galaxies, and use it as the primary distance indicator. and (2) then to use T-F, F-P, type Ia, and SBF as the secondary indicator to greater distance to determine H_0. They were able to achieve both goals. With the Cepheids calibration, (Table 1), they were able to much better calibrate the secondary indicators, the error on these indicates decreased from 10-20%, to 4-10%, because of that.

This plot (Figure 4) shows Hubble diagram, the diagram that shows the relation between distance and velocity, of key project results. Hubble diagram is a very important concept, which you should all remember. It plots D vs. v, and the slope of Hubble diagram is H_0. Also, when you go really far, it is going to curve, deviate from a straight line due to GR effect.

So this figure shows the Hubble diagram, with different indicators of Hubble constant, T-F, F-P, SBF, Ia and II (expanding photosphere). They are all consistent, no systematic differences among them. Table 12 shows the values and uncertainties. The final answer from the Key Project team is.

$$H_0 = 72 \pm 8\text{km s}^{-1}\text{Mpc}^{-1}.$$

However, this is not the end of the story.

2 Tension in H_0

However, this is not the end of the story. You might ask in this era of precision cosmology from WMAP and Planck, why do we still worry about Hubble constant from traditional method. The result is local distance measurements presented a completed independent way from CMB, which relies of growth of structure in the early universe.

It started with Adam Riess and colleagues (arXiv: 1604.01424) published the results from their most recent HST observations, in which they lowered the uncertainly of H_0 to 2.4%. The bulk of the improvement comes from new, near-IR observations of Cepheids in typeIa SN host galaxies. In 2016, the best estimate is $H_0 = 73.24 \pm 1.74$. This is 3.5-σ than the 2016 Planck of 66.93 ± 0.62.

The main improvement comes from (1) a much fewer step in the “distance ladder”. There are now three steps involved:

- Local geometric measurements (parallex!!) to Cepheids
- Cepheids to SN1a
- SN1a in Hubble flow to determine H_0.

(2) a much better determination of Cepheids distance in local galaxies with good Cepheids and SN1a, usually using IR relation.

(3) a larger number of paralllex with Cepheids, with GAIA and HST parallex. This is in the newest 2018 paper.

Figures show:

- History of H_0
- New distance ladder
- Error budget (Riess 2016 Figure 1)

- HST observations of SN1a hosts (Figure 2)
- Planck H(z) determination
- Riess+16 Figure 13

The very latest paper using GAIA DR3: $H_0 = 73.2 \pm 1.3$.

This is to be compared with Planck 2018 result: $H_0 = 67.27 \pm 0.60$ km s^{-1} Mpc^{-1}. This is a 4.2-σ result.

Riess et al. (2016) discussed possible solutions, other than observational issues. Changing most assumptions in our cosmological models do not help very much. However, An increase in the number of relativistic species in the early Universe increases the radiation density and expansion rate during the radiation- dominated era, shifting the epoch of matter-radiation equality to earlier times. The resulting reduction in size of the sound horizon (which is used as a standard ruler for the CMB and BAO) by a few percent for one additional species (N_{eff} = 4) increases H_0 by about about 10% for a flat Universe, more than enough to bridge the divide between the local and high-redshift scales. This N_{eff} is basically the number of relativistic species in the early universe (other than photon). We only now three species of neutrinos. If there is another species, which can be called "dark radiation", then we can reconcile those two.

We are still early in the days of this puzzle. But don't be surprised if this becomes serious discussion. Another example of cosmological as a tool for new physics.

3 Age of the Universe

Deceleration makes the universe younger than a Hubble time. Looking back in time, the expansion was faster and therefore we reach the Big Bang more quickly. *Draw pictures.*

For $\Omega = 1$ universe, the age of the universe is $2/3H_0$. Alternatively, $H_0 t_0 = 2/3$.

For open universes without lambda, $H_0 t_0$ can be larger. An empty universe $\Omega_m = 0$ has $H_0 t_0 = 1$; this is sometimes called a coasting universe. For reasonable $\Omega_m = 0.3$ or so, the result is about 0.85.

Positive Λ allows $H_0 t_0 > 1$. In fact, as $\Omega_m \to 0$, the Hubble parameter becomes time-independent, $\dot{R} = H_0 R$ means that $R \propto \exp(H_0 t)$, and we have an infinitely old universe. More on that later!

Lesson: Lower matter densities yield less deceleration and hence older universes for a fixed Hubble constant.

Observationally, the prejudice toward $\Omega = 1$ caused higher Hubble constants to be in conflict with age estimates of old stellar populations, in particular globular clusters. $H_0 = 50$ km/s/Mpc and $\Omega = 1$ gave an age of 13.3 Gyr. But $H_0 = 70$ would be only 10 Gyr. This was called the 'age crisis'.

Globular cluster ages: typically people say that the ages are 15 Gyr. Like many branches of astrophysics, errors bars have been quoted that in time turn out to have neglected important systematic effects. Opacities, abundance ratios, and the local distance scale have all caused these estimates to move around in the last 10 years.

One can also try to infer ages from white dwarf cooling.

With the currently favored Λ universes, the age of the universe is about 13–14 Gyr. My outsider's opinion is that this is within the errors on the stellar ages. I would also stress that it's remarkable that two completely disjoint programs agree to about 10%.

4 Ω_m

The debate on the value of Ω_m, the mass density, is a quite old one. The first attempt to measure Ω_m is to measure the deceleration of the universe to high redshift, using something called the brightest cluster galaxies, the brightest member of a cluster, which seems to be a quite good (10-15%) relative distance indicator. So what they were doing is really to measure $q_0 = 1/2\Omega_m$ if $\Lambda = 0$. The game started from as early as late 1950s, but it was never successful. The resulted q_0 tends to be all over the place, and it turned out that at that time, people didn't know how to model the evolution of galaxy luminosity itself, and at high redshift, the properties of BCGs changed, they are younger and brighter, so they are not good standard candles. In the end, this test didn't go anywhere. But such attempts unexpectedly grew into a whole new field, the field of galaxy evolution, when a group of young astronomers in mid 1970s began to question this method.

The second setback for measuring Ω_m is the realization that the universe is dark-matter dominated, and the visible, baryonic matter is but only a small fraction of it. So by counting stars and galaxies, it is very difficult then, to add up the total mass density of the universe, and get Ω_m in this way.

- $q_0 = \Omega_0/2$; brightest cluster galaxies;

- $\Omega_0 = 8\pi G\rho_0/3H_0^2$: dark matter??

Because of the difficulty of measuring Ω_m with deceleration, and with counting galaxies, for a long time, it is a rather confusing field. On one hand, people have known for a long time that baryonic matter is only a few percent, $\Omega_b < 0.05$. On the other hand, the nature of dark matter prevented people how to correctly estimate total Ω for a while. By 1980s, there are two obvious camps, one favors $\Omega = 1$, on the basis that it is the most elegant universe, inflation predicted it, flatness problem solved and so forth, the other favors $\Omega < 0.5$, as they just can't come up with enough dark matter in their measurement to close the universe. Actually, the majority astronomers who thought they knew the answer believed $\Omega_m = 1$. But in this case the minority won.

In the 1990s, the possibility of a low-mass universe gained substantial support through a number of very convincing observations, many of them independent. Although each observation has its strengths, weaknesses and assumptions, they all indicate that $\Omega_m < 1$.

Many of these methods that determined a low density universe involved using clusters of galaxies as the main tool. The idea is that since cluster of galaxies are the most massive and largest bound objects in the universe, they should have a very good mix of baryonic and dark matter, since the ratio of baryons to dark matter, or the mass to light ratio, would be close to the cosmic mean. And because they are bound, you can measure the cluster mass reasonably well, and use this to somehow derive the mean cosmic density. We will defer those to much latte because they require us to understand how cluster of galaxies are formed.

So by late 1990s, the overwhelming evidence is that Ω_m is small, we live in a light weighted, low density universe; but as I explained last time, theorists really want the universe to be flat, and with our cosmic triangle,

$$1 = \Omega_m + \Omega_\kappa + \Omega_\Lambda,$$

a low density universe can not be flat unless there is a cosmological constant. Is there? Now let's turn our attention to the second side of the cosmic triangle, the cosmological constant term, which determines the acceleration and the fate of the universe.

5 Dark Energy

You have heard from Jasmin about accelerating universe. What's the meaning of Λ? It is a replusive force in the universe that accelerates the universe. Looking at the energy equation of the Friedmann equation:

$$\ddot{R} = -\frac{4\pi G}{3} R \left(\rho + \frac{3p}{c^2}\right) + \frac{1}{3}\Lambda R$$

First, it causes the universe to accelerate. Second, it looks like some sort of energy term just as matter or photons... but there is nothing there. In elementary particle physics, it represents some sort of vacuum energy.

The story might stop right here with a happy ending– a complete physics model of the cosmic expansion– were it not for a chorus of complaints from the particle theorists. The standard model of particle physics has no natural place for a vacuum energy density of the modest magnitude required by the astrophysical data. The simplest estimates would predict a vacuum energy 10120 times greater.

(In supersymmetric models, it's "only" 1055 times greater.) So enormous a Λ would have engendered an acceleration so rapid that stars and galaxies could never have formed. Therefore it has long been assumed that there must be some underlying symmetry that precisely cancels the vacuum energy. Now, however, the supernova data appear to require that such a cancellation would have to leave a remainder of about one part in 10120. That degree of fine tuning is most unappealing.

The cosmological constant model requires yet another fine tuning. In the cosmic expansion, mass density becomes ever more dilute. Since the end of inflation, it has fallen by very many orders of magnitude. But the vacuum energy density ρ_Λ, a property of empty space itself, stays constant. It seems a remarkable and implausible coincidence that the mass density, just in the present epoch, is within a factor of 2 of the vacuum energy density.

Given these two fine-tuning coincidences, it seems likely that the standard model is missing some fundamental physics. Perhaps we need some new kind of accelerating energy– a "dark energy" that, unlike Λ, is not constant. with the goal of solving the coincidence problems.

The dark energy evinced by the accelerating cosmic expansion grants us almost no clues to its identity. Its tiny density and its feeble interactions presumably preclude identification in the laboratory. By construction, of course, it does affect the expansion rate of the universe, and different dark-energy models imply different expansion rates in different epochs. So we must hunt for the fingerprints of dark energy in the fine details of the history of cosmic expansion.

The wide-ranging theories of dark energy are often characterized by their equation-of-state parameter

$$w = p/\rho c^2,$$

the ratio of the dark energy's pressure to its energy density. The deceleration (or acceleration) of an expanding universe, given by the general relativistic energy equation:

$$\ddot{R} = -\frac{4\pi G}{3}R(\rho + 3p/c^2) + 1/3\Lambda R.$$

In case we don't have cosmological constant:

$$\ddot{R} = -\frac{4\pi G}{3}R(\rho + 3p/c^2) = -\frac{4\pi G}{3}\rho(1 + 3w).$$

So the expansion of the universe would depend on this ratio w. Thus the expansion accelerates whenever w is more negative than ?1/3, after one includes all matter, radiation, and dark-energy components of the cosmic energy budget.

Each of the components has its own w: negligible for nonrelativistic matter, obviously, for them, $p = 0$, pressureless. $w = +1/3$ for radiation and relativistic matter, for them $p = \rho/3c^2$. and $w =$?1 for Λ. Let's say how this is the case:

$$\Omega_\Lambda = \Lambda/3H^2,$$

We can write:

$$\Omega_\Lambda = 8\pi G\rho_\Lambda/3H^2,$$

similar to the way we write $\Omega_m = 8\pi G\rho/3H^2$. So $\Lambda = 8\pi G\rho_\Lambda$. Now for the energy equation, if we have only Λ term:

$$\ddot{R} = 1/3\Lambda R = 8/3\pi R\rho_\Lambda,$$

so comparing this with the equation above, we then have:

$$-2 = 1 + 3w \rightarrow w = -1.$$

That is, Λ exerts a peculiar negative pressure!

In fact, the energy equation tells us (see notes before):

$$\ddot{R} = -4\pi G/3\rho(1 + 3w),$$

$$dR/d\rho = -3\rho(1 + w),$$

$$\rho \propto R^{-3(1+w)}.$$

General relativity tells us that each component's energy density falls like $R^{-3(1+w)}$ as the cosmos expands. Therefore, radiation's contribution falls away first, so that nonrelativistic matter and dark energy now predominate. Given that the dark-energy density is now about twice the mass density, the only constraint on dark-energy models is that w must, at present, be more negative than -1/3 to make the cosmic expansion accelerate. However, most dark-energy alternatives to a cosmological constant have a w that changes over time. If we can learn more about the history of cosmic expansion, we can hope to discriminate among theories of dark energy by better determining w and its time dependence. Unfortunately, the differences between the expansion histories predicted by the current crop of dark-energy models are extremely small. Distinguishing among them will require measurements an order of magnitude more accurate than those shown in figure 3, and extending twice as far back in time. There are many telescope, both ground-based, and from space, that plan to observe many more distance SNe, to constrain the expansion of the universe much better, and to constrain the equation of state.

To summarize:

- $w = 0$: matter;
- $w = 1/3$: relativistic particles, photons, neutrinos;
- $w < -1/3$: accelerating universe;
- $w = -1$: cosmological constant

6 Cosmic Dynamics Experiment

We will discuss other dark energy experiments later on in this class. But I do want to mention one of the more ambitious ideas that I encountered. Issue with many DE experiments is that they suffer from systematics. SN tests you always worry about dust and evolution; other tests based on large scale structure have subtle effect with the growth of structures and the bias of tracers. People always wonder about what if one can directly measure the dynamical expansion of the universe, i.e., . first derivative of redshift, or, basically wait long enough to see the redshift of the object change with time. You can show (which is a homework problem) that:

$$dz/dt = (1+z)H_0 - H(z),$$

so for an object at rest, if you long enough, then because of the expansion of universe, the line will shift. If you observe this over a range of redshift, then you are mapping H(z), or Friedmann equation directly. The only problem is that this shift happens at Hubble time scale, i.e., dz/dt is extremely small. $\sim 10^{-10}$ per year, requiring velocity accuracy of cm/s, or about two orders of magnitude better than exoplanet surveys have achieved.

But that didn't stop people from planning. In fact, this is one of the science drivers for E-ELT....

Use Lyα forest observations. Very stable. 4000 hours on E-ELT over 20 years for one set of measurements. 10% of time available. Accuracy and model independent measurements of expansion history. Worth it?

Lecture Notes: Hot Big Bang
Feb 2024

1 CMB

We will now give a somewhat qualitative description of the early history of the universe, the generation of the CMB, BBN.

The cosmic microwave background (CMB) was discovered in 1965 by Penzias and Wilson. While the discovery was serendipitous, such a background had been theorized as early as 1948 on the basis of the effects it would have on nucleosynthesis. Big bang theory makes two fundamental predictions when Gamow first proposed it. First, actually the reason that big bang was proposed, was to explain, and to make predictions of the abundance of elements in stars, why stars do not consist of H only. This was the days before nuclear synthesis in stellar interior was worked out by Hans Bethe, and later by Burbidges, Hoyle, and Flower. So there was no explanation yet on why there are 25% of Helium in stellar atmosphere etc. It turned out that BBN can not explain the synthesis of elements beyond the lightest ones, and was not a successful theory. But it did make predictions on the primordial abundance pattern, and of course, on the relic photon field that we now know as CMB.

The CMB has three truly remarkable properties:

1. It is nearly perfectly isotropic.
2. It is a nearly perfect blackbody.
3. It is a lot of energy!

Moreover, it has not been resolved into discrete sources, even at arcsecond resolution.

These facts strongly favor an origin of the CMB that lies beyond the usual astrophysical processes that involve stars and galaxies. Energetically, the CMB dwarves the other known contributions to the intergalactic radiation field. The CMB is now known to have a temperature of 2.725 Kelvin which corresponds to an energy density of 4.2×10^{-13} ergs cm^{-3}. All other sources of extragalactic background light (e.g. the optical, X-ray, and submillimeter backgrounds) add up to $60 \pm 20\,\mathrm{nW\,m^2\,sr^{-1}}$, about 200 times less than the CMB ($12,600\,\mathrm{nW\,m^2\,sr^{-1}}$). So it is the dominant sources of photons in the universe.

this figure summarizes the amount of radiation background as a function of wavelength. νI_ν notation; origin of each components.

The nearly perfect blackbody spectrum of the CMB means that it must have been emitted from a very optically thick source. Since we can see radio and optical sources out to $z > 6$, the optical depth to that redshift must be small. Hence, the CMB was emitted prior to $z > 6$.

For a value of $\Omega_b h^2 = 0.02$ (explain this), the CMB today amounts to 1 MeV per baryon in the universe, and this increases as $1 + z$ at earlier times. Converting hydrogen to helium produces about 4.5 MeV per baryon, so even burning all of the hydrogen in the universe would only barely produce the CMB at modest redshift.

All these features point to the natural explanation that CMB came from the hot big bang, at redshift >> 1. If that were the case, then as we have derived from previous lectures, and this is something that you should know: for a relativistic specie, $p = \rho/3c^2$ as the equation of state, and the energy density goes as $\rho \sim R^{-4} \sim (1+z)^4$. Remember that the easy way to remember this is the number density goes as R^{-3} and the frequency of the photons go as R^{-1}. Since we have, for photons, $\rho \sim T^4$, that means the temperature goes as $T \sim 1/R \sim (1+z)$. So the tempertaure of the CMB would increase as $(1+z)$ to high redshift. Now if this can be observed, then it would be a very strong evidence that CMB really comes from high-redshift sources.

It is indeed observed. Example: work by Jian Ge, who was a graduate student here at Steward at the time, and now a professor at Florida building exoplanet instruments, and prof Jill Bechtold, using the MMT at Mt. Hopkins, measured the CI absorption line in bright high redshift quasars, which is caused by very cold gas clouds along the line of sight. You can see the different absorption lines caused by fine structure splitting of energy levels; and the relative strength of the lines are determined by the number of atoms in these fine structure states (with different J number, as L and S are coupled differently). The relative numbers are determined by the excitation temperature. In these very cool clouds, with no stars etc., the excitation temperature is thought to be the same as CMB temperature, which is the only heat source. So, by measuring these lines in high-z quasar, you can measure CMB temperature at high-z as well. To make the long story short, what these groups found was that the CMB T goes exactly as $T_0(1 + z)$ in these quasars.

2 Radiation dominated era

So we know that the matter density goes as $(1 + z)^3$, while the radiation density goes as $(1 + z)^4$. Then however small the CMB density, in term of mass density, is, there must be a time when the radiation were dominating, as it increases faster with redshift. When does it happen? Here we are looking for the time when the matter energy density equals the radiation density:

$$\frac{\rho_r}{\rho_m} = \frac{\sigma T^4}{\Omega_m (1+z)^3 c^2} = \frac{2.48 \times 10^{-5}(1+z)}{\Omega_m h^2}.$$

Here $H = 100h$km s^{-1}Mpc^{-1}. So at redshift $z > 4 \times 10^4 \Omega_m h^2 \sim 3000$, the universe was radiation-dominated, even though now it is matter-dominated. This is actually not quite complete yet. Here we are assuming that the radiation, or relativistic particle in the universe is only photons. It is not, there are also three kinds of neutrinos the same as photons. For reasons that we will discuss later this week, it turns out that there is also a neutrino background with a temperature of 1.95K. And adding that into the balance, the radiation dominated error lasted somewhat longer, to $z \sim 2.4 \times 10^4 \Omega_m h^2 \sim 1800$. We have derived before the dynamics in the radiation dominated era. Note that in this case, Ω is always one and cosmological constant is always negligible. So we have: $R \sim t^{1/2}$, or $T \sim t^{-1/2}$ as the universe cools down from the big bang.

The present-day photon-to-baryon radio is another key cosmological parameter. Assuming $T_0 = 2.725K$,

$$\frac{N_\gamma}{N_B} = \frac{3.6 \times 10^7}{\Omega_b h^2}.$$

Note that as far as we are not creating photons or baryons, this relation will hold to high redshift. Let's see how long it will hold. This is the same as measuring $\Omega_b h^2$, which we will talk about towards the end of this week.

3 Recombination Epoch

The hot Big Bang model explains the CMB as the relic radiation from the hot, early universe. At earlier times, the CMB photons were more energetic, with the temperature of the blackbody scaling as $1 + z$. At $z < 1000$, the photons are too cold to keep the hydrogen in the universe ionized, and so the optical depth is low. At $z > 1000$, the photons are hot enough to ionize hydrogen. This creates a high-optical depth plasma that can thermalize the spectrum. The transition from an ionized plasma to a neutral gas at $z \approx 1000$ is called "recombination".

At first, the fact that recombination occurs at $z \approx 1000$ may be surprising, since the CMB temperature at that epoch is only around 0.25 eV, far less than the 13.6 eV needed to ionize hydrogen. The temperature corresponds to 13.6eV is about 150,000 K. The discrepancy occurs because the number density of photons is so much larger than that of baryons—there are roughly 2 billion CMB photons per baryon—that even the very energetic tail of the Planck spectrum supplies enough photons to ionize the universe.

Now let's see how this works out: since we are only considering the highest energy photons, we are only worrying about the Wien, or exponential tail, of the black-body curve. So the total number of photons with $h\nu > E$ in the limit of $h\nu >> kT$ is:

$$n(>E) = \int_{E/h}^{\infty} \frac{8\pi\nu^2}{c^3} \frac{d\nu}{e^{h\nu/kT}} = \frac{1}{\pi^2}\left(\frac{2\pi kT}{hc}\right)^3 e^{-x}(x^2+2x+2),$$

where $x = h\nu/kT$. Now the total number density of photons in a BB spectrum at temperature T is:

$$N = 0.244\left(\frac{2\pi kT}{hc}\right)^3.$$

Therefore, the fraction of photons of the BB spectrum with energy greater than E is then:

$$\frac{n(>E)}{n} = \frac{e^{-x}(x^2+2x+2)}{0.244\pi^2}.$$

Roughly speaking, the gas will be ionized if there are as many photons with $h\nu > 13.6eV$ as there are hydrogen atoms, that is, we need only one photon in $3.6 \times 10^7/\Omega_b h^2$ of the photons in the CMB to have energy greater than 13.6eV to ionize the gas. And since the fraction above is determined by T, and $T \sim (1+z)$, that will determine the time when recombination happened. Now assume $\Omega_b h^2 = 0.04$, then this fraction is 10^{-9}, so then we will find: $x = E/kT = 26.5$, and the temperature needed at that time is: $T = 150,000/26.5 = 5600K$. Since the CMB temperature now is 2.7K, then the recombination happens at $z = 5600/2.7 \sim 2000$. Detailed calculations show that the pregalactic gas was 50% ionized at a redshift of $z_r \sim 1500$. This is the epoch of recombination. The universe was fully ionized before then, it was hot plasma, which recombined at $z \sim 1500$. Before the recombination epoch, the universe was full of electrons. It has very important consequences.

We know that electrons and photons interact, through the simple process of Thompson scattering. Without going to detail, which is described in Longair's book, we have the optical depth of Thompson scattering for a completely ionized plasma as:

$$\tau = 0.035\Omega_b/\Omega_m^{1/2} h z^{3/2}.$$

So for any reasonable value, $\tau >> 1$. Before the recombination era, the universe was opaque. Therefore, the universe beyond redshift of about 1100 is unobservable due to this huge optical depth. Any photons originating from larger redshift were scattered many times before they propagated to the Earth and all information about their origin is lost. It is called the photon barrier. Also, if there is no further scattering of photons of CMB, the redshift of about 1000

Before	$z \sim 1100$	After
$p + e^{-}$ plasma	recombination/phase transition	HI
$\tau >> 1$ opaque	last scattering/photon barrier	$\tau << 1$, transparent
$T_e = T_{CMB}$	decoupling	$T_e < T_{CMB}$ adiabatic expansion

becomes the last scattering surface. So what we are observing the CMB is really this last scattering surface, which can be regarded as the photosphere of the big bang. After the last scattering, the universe becomes optically thin to CMB photons, and they do not interact with matter anymore, in other words, electrons and photons are decoupled.

During the radiation era and before CMB era, we can no longer ignore the small energy change during the scattering process, i.e., the Compton effect. Because of the Compton effect, and the very large scattering optical depth, photon, or energy field, and matter are strongly coupled and they have the exact same temperature. The radiation dominated era leads us to the idea that the universe was much hotter in the past and that it was in fact gravitationally dominated by radiation.

Recombination is a landmark epoch in the history of the universe. It is the photon barrier, the last scattering surface for Thompson scattering, and when radiation and matter decoupled.

Recombination is an epoch, not an event, in the sense that it takes a period of time. We will show that later, but it is about $\Delta z = 200$.

One good way to look at the CMB is the analogy of stellar photosphere. In this sense, recombination is the photosphere of the universe, when optical depth is one. The other good reason for this analogy is that CMB spectrum is similar to stellar photosphere spectrum as well. It is a black body continuum plus recombination lines. More on that later. Since there is no other opacity source, because no matter, the spectrum is far simpler. But one does need to model recombination.

One more note before we look at what is going on in the matter dominated era. If the CMB photons last scattered at some high redshift, e.g. 1000, then photons that arrive to us from very different directions were emitted by regions in the universe that were apparently out of causal contact. The age of the universe at CMB era is about 1 million years. So size of the universe that could have causal contact, the signal could travel is $c \times t$ which is about about 0.3 Mpc, which has an angular diameter of about 4 deg on the sky. All the patches on the sky with separation larger than this are out of causal contact at the recombination time. If these patches were actually out of causal contact, then there would be no way for a causal process to arrange that the CMB in those patches would have the same temperature.

Hence, the near isotropy of the CMB is an amazing fact about the universe, implying that the causal structure that we extrapolate from the matter that we know about *must be wrong.* Inflation provides a striking solution to this problem.

4 Early Epochs

Before recombination and during radiation dominated era, matter and radiation are strongly coupled by Compton scattering. In this case, the electrons and photons are transferring energy with each other during the collisions, and matter will have the same temperature as photons, or CMBs.

The thermal history of earlier universe is well described by simple thermal dynamics and particle physics. Let me outline briefly here (Figure):

(1) we can extrapolate back to redshift of about 3×10^8 when the radiation temperature was about $10^9 K$. This temperature is sufficiently high for the background photons to have γ-ray energies of about 100 keV. In this case, the high energy photons in the tail of the Planck distribution will have enough energy to dissociate light nuclei such as helium and D. At earlier epochs, all nuclei are dissociated in protons and neutrons. Now we are travelng back in time. To travel forward in time, this is the epoch when the process of primordial nucleosynthesis took place, we will discuss that in just a minute.

(2) At $z \sim 10^9$, the energy of the photon field is about half Mev. This is the energy of electron-positron annihilation. So the e-p pari production from the thermal background can take place and the universe was then full of e-p pairs. So now if we are travelng forward, the e and p are annihilating at this epoch, annihilate, their energy is transferred to the radiation field. THis is the little discontinuity in the figure.

(3) At slightly earlier epoch, the opacity of the universe for weak interactions became unity. We know that weak interactions involve neutrinos, in the same way that E&M interaction involves photons. So this epoch is the epoch of neutrino barrier, since to the photon barrier at $z \sim 1000$. At this epoch, neutrinos become uncoupled.

(4) Extrapolating to about $z \sim 10^{12}$ when the temperature of the radiation field was high enough for baryon-antibaryon pair production. Just as in the case of the epoch of e-p pair-production, the universe before then was flooded with baryons and anti-baryons. Therefore, there is any a slight discontinuity in temperature. Here there is one of the cosmological problems, namely the baryon asymmetry problem. In order to produce the matter-dominated universe we live in today, there must have been a tiny asymmetry between matter and antimatter in the very

early universe, about one in 1 billion excess of matter. This asymmetry is the reason for the photon/baryon ratio we have today. It must originate from the very early universe.

(5) Finally, we can this to as far as we trust our particle physics, which is probably at least believable to about 100 GeV. How far back more one can push is then a matter of taste. Some have no hesitation to go all the way to the Planck era, when particle physics and GR merge and the physics as we know become irrelevant. Note that Planck scale is defined as when Schwardschild scale, $\lambda = Gm/c^2$, or the size of horizon, is the same as the Compton wavelength, or the size determined by uncertainty principles: $\lambda = h/(mc)$. Combining these two, we have:

$$\lambda_p = (Gh/c^3)^{1/2},$$

and Planck time:

$$t_p = (Gh/c^5)^{1/2} = 1.35 \times 10^{-43} sec.$$

Beyond this, our view of space-time has to be changed. We are entering the regime of string theory, which is beyond this course.

Most part of these early history of the universe will be beyond our class, which, as we noted in our first lecture, is a cosmology class from an extragalactic viewpoint. So we will not discuss in detail things that are really beyond our current means of astronomical observations. But how far can we push? We talked about photon barrier, that is, at the CMB era, $z \sim 1200$, the optical depth due to Thompson scattering is large. CMB acts as a photosphere of the universe, so we can't see through it. It is a perfect BB, with no more information other than a temperature. How do we know, or test our model beyond that? There are a few ways that early universe left imprints on later era observations.

- CMB fluctuations. The T fluctuation of CMB is a result of early physical processes, in particular, inflation. We will discuss this in a few weeks.
- g-wave. We don't know how to do that. But in principle. Also, it has consequences on CMB structure.
- Neutrino. As we discussed, the neutrino barrier is at much higher redshift, $z \sim 10^{10}$ or so. It left us with a neutrino background at $T = 1.95K$, a temperature we will justify later. So if we can observe ν background as we do with photons, maybe. But we don't know how to do that either.
- BBN. What the last point says is that the signatures of weak interaction (and strong interaction) in the universe are preserved at higher redshift. We can observe phenomena related to weak and strong interaction, i.e., nuclear interaction, at higher redshift to

test our physics. The key result of such interactions is the synthesis of heavy element in the universe a process similar to the nuclear reaction that we are familiar with in the interior of stars, with one major difference. In stars, we think about nucleosynthesis as a process with an increasing T. In BBN, temperature are decreasing rapidly as the universe expands. So your nuclear reaction rate is always competing with the Hubble expansion. When it is longer than Hubble time, it is not relevant anymore. So during BBN, only a small number of light elements formed. The abundance of these element provides a sensitive test to our cosmology, especially to the number density of baryons.

5 Equilibrium Abundances in the Early Universe

Before we discuss BBN, we should consider a bit further the role of neutrinos and weak interaction in the early universe. I hope you still remember some of your particle physics and statistical physics stuff.

The CMB and big bang leads us to the idea that the universe was much hotter in the past and that it was in fact gravitationally dominated by radiation. It has important consequence to the species of particles in early universe.

Imagine what would happen if the universe were very hot. For example, let's consider that the temperature of the CMB were 10 MeV/k. Two photons colliding could produce an electron-positron pair.

$$e^+ + e^- \leftrightarrow \gamma + \gamma$$

How frequent is such a reaction? A given photon will have a reaction rate $nc\sigma$, where σ is the cross-section of the reaction. The cross-section is approximately the Thompson cross-section. In detail, it's about $(3/16)\sigma_T(m_e c^2/E)^2$, which is about 3×10^{-28} cm^2 at 10 MeV.

The density of photons today is 411 cm^{-3} at a temperature of 0.00023 eV. The redshift to make the CMB be at 10 MeV is 4×10^{10}. Hence, the density of photons is then 2.5×10^{34} cm^{-3}. So the reaction rate is $2 \times 10^{17} s^{-1}$.

What is the age of the universe at this redshift? The Hubble constant will be

$$H^2 = \frac{8\pi G\rho}{3}$$

At these redshifts, the density of the CMB is far higher than that of the protons, neutrons, and other matter. Using just the density of the CMB, we would have $\rho c^2 = aT^4$ where $a = 7.56 \times 10^{-15}$ ergs cm^{-3} K^{-4}. We had 4.2×10^{-13} ergs cm^{-3} at $z = 0$; here we have

$(1+z)^4$ more, 1.1×10^{30} ergs cm^{-3}! Dividing by c^2 gives a mass density of 1.2×10^9 g cm^{-3}. That makes $H = 25s^{-1}$. The age of the universe is of order $H(z)^{-1}$, which works out to 0.04 seconds.

Hence, a photon can interact of order 10^{16} times in the age of the universe. This means that the number of photons and electrons will come to a statistical equilibrium. How many of each should be around?

For weakly interacting gas of particles in thermodynamic equilibrium, the number density of particles is

$$n = g \int \frac{d^3p}{h^3} f(E)$$

where g is the number of spin states and $f(E)$ is the occupation number of the non-interacting states. Statistical mechanics says that the occupation of a given state is

$$f(E) = \frac{1}{\exp[(E-\mu)/kT] \pm 1}$$

for fermions and bosons. Here, $E^2 = m^2c^4 + p^2c^2$ and μ is the chemical potential of the species.

The energy density is

$$u = \rho c^2 = g \int \frac{d^3p}{h^3} f(E) E$$

We can switch variables to E by using $d^3p = 4\pi\sqrt{E^2 - m^2c^4}\, E\, dE/c^3$. Then

$$\rho c^2 = \frac{4\pi g}{c^3h^3} \int dE \frac{E^2\sqrt{E^2 - m^2c^4}}{\exp[(E-\mu)/kT] \pm 1}$$

For a relativistic species ($m = 0$) with $\mu = 0$, the energy density integral is

$$\rho c^2 = \frac{4\pi^5 k^4}{15c^3h^3} g T^4 = \frac{g}{2} a T^4$$

for bosons. You know this result from the Planck spectrum. For fermions, there is a reduction by a factor of 7/8.

The number density of particles is

$$n = \frac{8\pi k^3}{c^3h^3} \zeta(3) g T^3 \propto T^3$$

for bosons. For fermions, there is a reduction by a factor of 3/4.

We are dealing with three different kind of particles in the early universe:

- photons. They are massless bosons with g=2.

$$N = 0.244(2\pi kT/hc)^3 m^{-3}, E = aT^4;$$

- nucleons, electrons and their antiparticles. They are fermions with g=2.

$$N = N^- = 0.183(2\pi kT/hc)^3 m^{-3}, E = 7/8aT^4;$$

- neutrinos and antineutrinos. We are aware of three kinds of neutrinos. They are fermions with g=1.

$$N = N^- = 0.091(2\pi kT/hc)^3 m^{-3}, E = 7/16aT^4.$$

Hence, we would predict that the number of electrons should be 7/8 of the number of photons. This is vastly more electrons than we observe in the universe today. What happened to them?

As the universe cools, the temperature drops below the rest mass energy of the electrons. The annihilation rate of electrons and positrons became faster than the production rate. The electrons and positrons disappear in favor of photons.

The energy in the electrons and positrons (which was 7/4 of that in the CMB) becomes extra energy in photons. But this means that the photons have too much energy for a blackbody of their temperature. As the photons scatter off the residual charge in the universe, they can share their energy and achieve a new blackbody temperature. If the annihilations were faster than the redistribution, then the change in temperature would be at constant energy, which would mean that T would increase by $(11/4)^{1/4}$. However, in fact the annihilations are slower than the redistribution, which means that the change is accomplished adiabatically, i.e. at constant entropy. The entropy of a relativistic distribution goes $(\rho + p)/T = 4\rho/3T$, which scales as T^3. Hence, the temperature actually increases by $(11/4)^{1/3}$.

Note that this breaks the relation between R and T. As the universe expands, normally $T \sim R^{-1}$, but as the electrons and positrons are annihilating, the temperature drops less quickly.

Of course, as the temperature drops, the equilibrium density of electrons would go to zero. We apparently would have no electrons in the universe! Two effects intervene.

First, the annihilation rate depends on the density of electrons and positrons and eventually this becomes so small that the electrons can't annihilate in the age of the universe. The resulting abundance is known as the "freeze-out" abundance. Clearly, the result depends on the annihilation cross-section. Larger cross-sections make the final abundance smaller.

However, for electrons the freeze-out abundance would be far too small (moreover, there'd be a lot of positrons around!). Instead, the residual electron density is set by a tiny asymmetry (10^{-9}) in the number of electrons relative to the number of positrons. All the positrons annihilate on electrons, but the asymmetry is left over.

This is known as the lepton asymmetry. We don't know where it comes from!

In the early universe, the relativistic species determine the density. Any species with $kT > mc^2$ will have a number density similar to the CMB. We can count them!

$$\rho c^2 = (g_{boson} + \frac{7}{8} g_{fermi}) \frac{a}{2} T^4$$

In detail, we should allow for each species to have a separate temperature. We define

$$g_* = \Sigma_{boson} g_i \left(T_i/T_{cmb}\right)^4 + \frac{7}{8} \Sigma_{boson} g_i \left(T_i/T_{cmb}\right)^4$$

and then write $\rho c^2 = g_* (a/2) T_{cmb}^4$.

Then we have the Hubble constant, which is $H^2 = 8\pi G\rho/3$. This gives $H \propto g_*^{1/2} T^2$, which means that the age of the universe is roughly $t \sim T_{MeV}^{-2}$ seconds.

As the universe cools, species for which the rest mass is important annihilate and become trace constituents.

6 neutrinos in early universe

Next, consider neutrinos. These have too slow a reaction; they stop scattering or annihilating at 1 MeV, so their number density is frozen, regardless of whether they are relativistic today.

The cross-section of the neutrinos are weak, typically $G_F^2 T^2 = 5 \times 10^{-44} T_{MeV}^2$ cm^2.

The redshift to achieve 1 MeV temperatures is roughly 4×10^9 (but this is actually a bit too high by $(4/11)^{1/3}$), which makes the density 2.5×10^{31} cm^{-3}. The interaction rate is $\Gamma = n\sigma c = 0.04\ s^{-1}$. This is comparable to the Hubble constant, so we have roughly 1 encounter per Hubble time (this came out to 25, but we dropped lots of factors).

Above 1 MeV, there are plenty of interactions to make neutrinos and antineutrinos. We get full blackbody populations.

However, the neutrinos stop interacting at $T < 1$ MeV. This is known as decoupling. These backgrounds of 3 species of neutrino and antineutrino should still be around in the universe! There should be roughly as many low-energy neutrinos as there are photons!

This cosmic neutrino background has never been detected directly! The neutrinos have a tiny energy and no one has thought of a way to detect them directly. However, we can infer their presence cosmologically, by their effect on nucleosynthesis in the early universe.

However, as a minor detail, the early decoupling of the neutrinos means that they do not share in the energy released by the annihilation of the electrons and positrons. Hence, the photon temperature rises by $(11/4)^{1/3}$ relative to the neutrinos. In other words, the neutrinos have a temperature of $(4/11)^{1/3} \times 2.725$ Kelvin, which is 1.95 K. This means that there are fewer neutrinos than photons (each spin state is down by a factor of $(3/4) \times (4/11)$ relative to that of photons, so we have $6 \times (3/4) \times (4/11) \times 411/2 = 336$ neutrinos per cubic centimeter.

When the neutrino's decouple, they are very relativistic and hence have a momentum distribution given by a (fermionic) blackbody. However, we think the neutrino does have a small mass. Today, this mass is considerably higher than the kinetic energy. This can make a significant mass contribution to the universe! If a particular species of neutrino had a mass of about 30 eV, the 112 neutrinos per cubic centimeter would suffice to make $\Omega = 1$!

Hence, we can infer that the neutrino (if stable) cannot have a mass between about 30 eV and 1 MeV! Otherwise, the universe would have $\Omega \gg 1$, in conflict with observations. In fact, we think the limit can be placed lower, but we'll get to that later.

One of the remarkable properties of the Hot Big Bang is that standard particle and nuclear physics predicts the formation of certain elements in the early universe. Indeed, the large amount of helium and simple existence of deuterium would be very hard to explain were it not for the Big Bang. It is a remarkable piece of corroborating evidence for the Hot Big Bang that the predicted abundances work out to be close to those observed.

At very high energies, $T \gg 1$ GeV, there are lots of quarks and antiquarks around. As the temperature drops to a few MeV, the quarks and antiquarks annihilate, leaving only a small (and unexplained) residual population of quarks.

These quarks are bound into protons and neutrons. The neutron is slightly heavier than the proton, by about 1.293 MeV. Neutrons and protons can transmute by weak interactions

$$n + \nu_e \leftrightarrow p + e^-$$

If these reactions were fast (compared, as usual, to the Hubble constant), then the relative density of the two would be

$$\left[\frac{n}{p}\right] = \exp\left(-\frac{1.3\text{ MeV}}{kT}\right)$$

As the temperature dropped well below an MeV, all the neutrons would disappear and the universe would be left with only hydrogen.

Or would it? As you know from stellar physics, there are heavier elements that are energetically favored. If we really had equilibrium all the way to low temperatures, these nuclei would be favored and the universe would be made of iron!

Both of these predictions are wrong because the interaction rates are not arbitrarily fast. The process of nucleosynthesis is guided through a few particular channels that yield the standard predictions.

7 The decoupling of neutrinos and neutrino barrier

First, let's consider the weak interaction that governs the conversion of protons and neutrons. We saw before that two neutrino interactions froze out at about 1 MeV. We get a similar answer here.

Roughly, the time-scale for weak interaction is

$$t_{weak} = (\sigma N c)^{-1}$$

where σ is the cross section of weak interactions, $\sigma \propto E^2 \propto T^2$, and N is the number density $N \sim R^{-3} \sim T^3$, so $t_{weak} \sim T^{-5}$. And the age of the universe $t_H \sim R^2 \sim T^{-2}$. It is important to remember that the cross-sections fall sufficiently quickly with energy that the ratio of the interaction rates to the Hubble constant is a steep power of temperature, T^3. So the reactions really do shut off.

Detailed calculations find that the ratio of neutrons to protons drops to about 1:6 as the temperature drops below 1 MeV, corresponds to $T \sim 10^{10} K$. Remember that the age of the universe is about 1 second.

A by-product of this is that we have also derived the time when the universe was transparent to neutrinos. Analogy to the photon barrier at $z \sim 1200$, we have a neutrino barrier, and thus a neutrino background. We can in principle use neutrino background to probe the universe when it was 1 sec old!

8 The Synthesis of the light elements

We now have 1 neutron and 6 protons in a cooling universe. If nothing else happened, then the neutrons would decay back into protons with a half life of 887 seconds. Remember that free neutrons are unstable!

However, the neutrons are saved. After a few minutes, the temperature of the universe is down to about 100 keV. At these temperatures, ^4He is favored. Most of the neutrons become bound into the Helium.

There are two questions that arise here:

1) In fact, ^4He is favored even at slightly higher temperatures (say 300 keV). Why doesn't form a little earlier?

2) Why, at some lower temperature, doesn't the helium fuse into ^{12}C and so on into heavier elements?

The answers are that the reaction pathways have bottlenecks. To form the helium, we have to fuse 4 protons. 4-body interactions are very rare; instead, nature prefers to link together 2-body interactions. The relevant one in this case involves deuterium, $n + p \to D + \gamma$. The deuterium then combines with another p or n to form ^3He or ^3H, and then one last nucleon to reach ^4He. Alas, deuterium is rather fragile and there are a billion times more photons than nucleons. So at hotter temperatures, the abundance of deuterium rises slowly. The production of ^4He is starved waiting for it!

Why stop with helium? The same reason that helium on earth doesn't fuse: the Coulomb barrier. As the universe cools, heavier elements are energetically favored, but the reaction rates are very slow because of the Coulomb repulsion of the nuclei. The delay in the formation of Helium helps here, as does the lack of a tightly bound nucleon with mass 5 or 8 (there's no good intermediate point between ^4He and ^{12}C).

So the process ends with most of the neutrons bound into ^4He. How much helium? In the delay of a few minutes to cool to 100 keV, the neutron fraction decays from 1:6 to 1:7. This means that there are 2 neutrons for every 14 protons. That makes 1 helium and 12 hydrogen. Helium is about 25% by mass ($Y = 0.25$).

This is very close to what is observed!

To a part in 10^4, all of the neutrons end up in ^4He, but the calculation of the exact neutron-proton ratio is actually rather detailed.

The neutron-proton ratio depends sharply on temperature as the weak interactions are freezing out, so these reactions must be modeled carefully.

Moreover, at about this time, the neutrino-neutrino interactions are slowing, so the neutrinos aren't necessarily held into a thermal distribution.

Moreover, the electron-positron pairs are annihilating, so the universe is getting reheated. It spends longer at certain temperatures than one would expect.

So, the actual freeze-out abundance present at $t \sim 100$ seconds requires a lot of detailed calculation. However, the principles are generally straight-forward: we know how to calculate electroweak cross-sections. Theoretical predictions are thought to be good to better than

1%.

9 Dependence on cosmological parameters

What dependences does the result have?

For a long time, the uncertainty in the overall amplitude of the weak interaction rates dominated the uncertainties. Weaker interactions mean earlier freeze-out's and higher n/p. However, this parameter is now well-measured.

If one increases g_* by adding extra relativistic species to the universe, then one gets faster expansion at a given temperature. This means that reactions freeze-out a bit earlier ($\Gamma = H$). In detail, $T_{fo} \propto g_*^{1/6}$. The result is higher n/p and more helium.

If the baryon-to-photon ratio is larger, then deuterium forms more easily. This makes ^4He production a bit earlier. However, this is a weak effect on ^4He. Factors of 10 in the baryon density make 3% effects on helium.

So, one reaches the remarkable conclusion that the abundance of helium depends primarily on the number of relativistic species in the universe! The argument has generally been that BBNS rules out a 4th neutrino species (even a sterile species)! The best fit value so far has $n_\nu = 2.3$ with $n_\nu = 3$ consistent at about 1-σ level. This constraint is quite bit earlier than the experimental physics constraint and it is truly remarkable.

Figure, Longair 10.2

$\eta = 10^{10} n_B/n_\gamma = 274\Omega_b h^2$

10 Observations

10.1 Helium-4

Observationally, the fraction of ^4He is close to what's predicted, but different groups do disagree at the 2% level. In any case, stars also make ^4He. Plotting ^4He against metallicity in stars or HII regions reveals a small slope (solar stars are ~ 0.29). The game is to assume that helium enrichment and metal enrichment are related and to try to extrapolate to zero metals. This isn't so bad, but the measurement of helium abundances in stellar atmospheres or the ISM is never perfect. Many estimates of the primordial helium are a bit lower than the BBNS prediction ($Y_p = 0.238 \pm 0.007$), but the uncertainties seem large enough to be consistent. Nevertheless, it is a triumph that the helium predicted from the early universe is a close match to that observed in metal-poor stars!

figure 10.3 of Longair

10.2 Deuterium and He-3

Trace amounts of other nuclei are also produced. Deuterium and ^3He are made on the way to ^4He, but as the temperature drops the Coulomb barriers cause the final reactions to freeze-out. This leaves a small amount of D and ^3He around (and ^3H, but that decays). The levels are a few parts in 10^5.

This production is *very* sensitive to the baryon density. The reaction rates are proportional to the density, so higher density means faster reactions, which means more complete burning to ^4He and lower D,^3He abundances.

If we could measure the "primordial" (i.e. pre-stellar) fraction of D and/or ^3He we would have a measurement of the baryon density!

D is relatively fragile. All known and common astrophysical processes destroy it. In particular, it is burned easily in stars as part of the pre-main-sequence evolution. Conservatively, one takes the observed D abundance as a lower limit. This means that it is an upper limit on the baryon density! In terms of the critical density, this limit is $\Omega_b h^2 < 0.025$. With $h > 0.5$, that means $\Omega_b < 0.1$ (and probably a fair bit less).

We observe the density of the local universe to be higher than this, typically Ω_m at least 0.2. This is strong evidence for non-baryonic dark matter!

It has always been dicey to argue about whether the D/H ratio in the Milky Way is primordial. Recently, we have gotten measurements at high redshift from absorption in QSO spectra. This argues for $D/H \approx 3 \times 10^{-5}$. If this is primordial, and it may well be, then $\Omega_b h^2 = 0.022 \pm 0.002$! We have a 10% measure of the baryon density! Measurement of deuterium in quasar absorption line is generally regarded as primordial. But because of the low D abundance, the observation is highly difficult. It is a weak transition next to H, because of the slight different in the reduced mass when having atomic number of 2 instead of one. So on the blue wing of Lyα line. It requires: (1) accurate measurement of H, which is always saturated in order to see D; (2) removal of contaminant in Ly alpha forest.

Local D/H measures are generally lower than this value, around 1.5×10^{-5}, and display significant scatter. It is thought that much of the deuterium in the ISM has been destroyed by cycling through stars.

This baryon density does prefer a relatively high value for ^4He, around $Y = 0.245$ or 0.25. This is a bit of a stretch.

Figure 10.4 of Longair

For a few years, there were claims of very high D/H ratios, leading to low baryon densities. I think these are all defunct now.

^{3}He is observed in oldest meteorites which is regarded as the value for solar nebula. The interpretation is complicated. D burns into He-3 which burns in He-4.

10.3 Lithium-7

^7Li is also produced in BBNS. It is produced by combining ^{4}He with ^{3}H at low baryon density and with ^{3}He at high baryon density. This produces a minimum abundance near where the predicted baryon density is.

Figure 10.5 of Longair

The predicted abundance is about 3×10^{-10}. This is a factor of 2 higher than what is observed in stars $(1 - 2 \times 10^{-10})$. The destruction of Lithium in stars is a fairly active and contested field. Lithium burns fairly easily in stars. Convective stars will destroy most of their surface lithium. But it can also be enhanced by collisions of cosmic ray protons and cold ISM gas. Figure shows a lithium platoon at low metallicity.

What makes the heavier elements? Everything above Carbon is made in stars. Most of the B and Be are made when cosmic rays hit heavy nuclei and split them in two. This is called spallation. Note that the spallation abundances are far below that of D; it is implausible that deuterium is produced by spallation.

Review by Olive astro-ph/0202486. Note that much of the BBNS literature uses the baryon-to-photon ratio $\eta = 5.36 \times 10^{-10}(\Omega_b h^2/0.02)$.

Most of the leverage in BBNS currently comes from D/H, as this puts an upper bound on $\Omega_b h^2$ that is very similar to the results from the CMB. Indeed, the latest CMB results push the baryon slightly higher, which would require a slightly lower D/H. Also, ^{4}He and ^{7}Li don't fit too well with that value.

BBN is one of the sometimes overlooked triumph of big bang. It is a highly specialized field. But this overlook might be just due to the simplicity and success of the whole prediction. I regard the BBN constraints on things like the number of relativistic species some of the most beautiful physics I encountered. It is the highest redshift direct probe of the universe up to now, with the exception of CMB constraints on inflation which is indirect and highly model dependent. With the BBN, we are safely trust our basic understanding of the expansion and thermal physics in early universe up to $z = 10^{10}$. Actually, things are most predictable at between photon barrier and neutrino barrier, at redshift between thousand and billion. Before that, we are more and more dependent on particle physics at the highest energy. At lower redshift, the inhomogeneity of the universe becomes important which results in the growth of structures, which we occupy our discussion for the rest of the semester.

ASTR 541 Lecture Notes: Perturbations and Large Scale Structure, 2024

1 Plan for the next month

So far, we've studied homogeneous cosmology. Our universe has had the same density at all points in space and have merely evolved in time. This has allowed us to make statements about the thermal history and fate of the universe. However, if we want to study objects like galaxies, we have to consider departures from **homogeneity.**

Galaxies and clusters of galaxies are complex systems, but the aim of cosmologist is not to explain all their details - that's the goal of galaxy class, astrophysicist. Here we seek to explain the origin of large scale structure of the universe, and how does it develop in later universe. We define the density enhancement $\delta\rho$, and the density contrast $\Delta = \delta\rho/\rho$. When $\delta << 1$, these perturbations grow linearly (actually it is linearly with scale factor in standard models). However, when these amplitude approaches unity, their subsequent development becomes non-linear and they rapidly evolve into bound objects, such as galaxies and clusters of galaxies, when many non-linear astrophysical effects become important, such as star formation and feedback. Our second part of this course is to deal with mostly small, linear perturbations and the observations of large scale structure, and how to use them to probe our cosmological model. Here our objectives are twofold, (1) to understand how density perturbations evolve in an expanding universe, and (2) to derive and account for the initial conditions necessary for the formation of the structures we observe.

- **perturbation $\delta(t)$, and cosmological tests**
- **given cosmology and observations, initial condition**

In the first few lectures of this part of the course, we will mostly due with the first question. We will go over how density and velocity perturbations grow in the universe, and then introduce statistics of large scale structure and the means to observe them at low redshift. Then we will treat the fluctuations of the CMB as a special case, which is especially important in understanding some of the initial condition issues and its impact in our understanding of the early universe.

Before diving into the math, some back-of-the-envelop considerations of the scale that we are talking about. You can simply calculate the mass density of a galaxy ($\sim$ 10 kpc), a cluster of galaxy ($\sim$ 1 Mpc), and then a supercluster, or large scale structure you see in redshift surveys ($\sim$ 10 Mpc). Comparing that to the mass density of the universe, e.g, for $\Omega = 0.3$. Turns out that $\Delta = \delta\rho/\rho$ is 10^6, 10^3 and a few, respectively. Note that for a bound object, such as a galaxy, $\delta\rho$, its density will not change with the expansion of the universe: it is a bound object, which means that the local gravitational field has already overcome the expansion and it is viralized and doesn't grow or collapse anymore (to the zeroth order, it doesn't change subtly). This is an important point: when an object has already gone non-linear, our Friedmann universe model doesn't work anymore. A bound, collapsed object is decoupled from the expansion of the universe and dominated by the local gravitational force. We will show that the minimum density of a viralized object is of order 200 the cosmic mean, so it is no longer subject to Hubble expansion, and one can safely ignore the GR effect on cosmological scale when dealing with individual objects at cluster scale or smaller.

Since $\rho \sim (1+z)^3$, that means that for galaxies, we will reach $\Delta \sim 1$ at $z \sim 100$ or so if nothing else happens. This immediately gives you that we should begin to consider galaxy formation from redshift of a hundred or lower, and clusters at $z < 10$. Turns out that most galaxies started to form at even lower redshift but not by more than an order of magnitude. So this gives you the basic scale of the redshift range that we care about.

2 Growth of small perturbations in the expanding universe

The universe is obviously lumpy on small scales, and we have argued that it gets smoother on large scales. This is the justification for considering the inhomogeneity as a perturbation to the homogeneous solution.

We have already done this kind of analysis once in the case of the ISM. We found a Jeans instability, in which perturbations grew exponentially if they had a longer wavelength than the Jeans wavelength and were stabilized by pressure otherwise.

Now we'll repeat this analysis in the case of an expanding universe. Two key differences: we need to deal with an expanding universe, and we need to use comoving coordinate system to simplify expressions.

But let's start with the basic hydro equations

$$\frac{d_c\rho}{d_ct} + \rho\nabla_x \cdot \mathbf{v} = 0$$

$$\frac{d_c\mathbf{v}}{d_ct} = -\frac{1}{\rho}\nabla_x p - \nabla_x\Phi$$

$$\nabla_x^2\Phi = 4\pi G\rho$$

where the $d_c/d_c t$ notation indicates a Lagrangian derivative and the ∇_x are derivatives with respect to the physical coordinate $\mathbf{x}$. This is the notation in fluid dynamics with a coordinate system that motion of a particular fluid element is followed. The first equation is the continuity equation, the second is the equation of motion, or Euler equation, and the third describes the gravitational potential, the Poisson equation.

You might be more familiar with the other coordinate system in fluid dynamics, the Euler system, in which we apply partial derivatives to qualities as a function of a fixed grid of coordinates,

$$\frac{d_c}{d_c t} = \frac{\partial}{\partial t} + (\mathbf{v} \cdot \nabla_x).$$

So in real calculations, such as when you carry out cosmological simulations, there are advantages and disadvantages of using either Euler or Lagrangian systems when setting up your calculations. We are dealing with cosmology, and with particles that are almost fixed in Hubble flow, so one more complication is that we want to express things in comoving coordinate. In particular, we define $\mathbf{x} = R\mathbf{r}$ where $R(t)$ is the usual expansion factor. This means $\nabla_r = R\nabla_x$.

We then split the velocity into a Hubble expansion term and a peculiar velocity term, $\mathbf{v} = H\mathbf{x} + \delta\mathbf{v} = H\mathbf{x} + R\mathbf{u}$. $\mathbf{u}$ is the comoving peculiar velocity, i.e. how far a particle moves in comoving coordinates per unit time.

$\delta\mathbf{v}$ is called a peculiar velocity. It's the velocity difference relative to the expectation of the Hubble law.

We will then switch from a Lagrangian derivative in physical units to a derivative in the comoving frame. This means that

$$\frac{d_c}{d_c t} = \frac{\partial}{\partial t} + (\mathbf{v} \cdot \nabla_x) = \frac{\partial}{\partial t} + (H\mathbf{x} \cdot \nabla_x) + (\delta\mathbf{v} \cdot \nabla_x) = \frac{d}{dt} + (\mathbf{u} \cdot \nabla_r),$$

where d/dt is differential over the comoving grids.

Finally, we will find it convenient to measure densities in fractional units relative to the background density $\rho_h(t)$. So $\rho = (1+\delta)\rho_h$.

We now can insert these substitutions and remove the homogeneous part of the solution.

Let's see how our hydro equations become in the next lecture.

For the continuity equation, we get

$$\frac{d\rho}{dt} + (\mathbf{u} \cdot \nabla_r)\rho + \rho\nabla_x \cdot (H\mathbf{x} + R\mathbf{u}) = 0$$

$$\rho_h \frac{d\delta}{dt} + (1+\delta)\frac{d\rho_h}{dt} + \rho_h(\mathbf{u} \cdot \nabla_r)\delta + (1+\delta)\rho_h(3H + \nabla_r \cdot \mathbf{u}) = 0$$

$$\frac{d\rho_h}{dt} = \frac{d(\rho(0)R^{-3})}{dt} = -3\rho_0 R^{-4} dR/dr = -3H\rho_h$$

Note that $\nabla_x \cdot \mathbf{x} = 3$. This means that the homogeneous terms cancel, leaving us with

$$\frac{d\delta}{dt} + (\mathbf{u}\cdot\nabla_r)\delta + (1+\delta)\nabla_r\cdot\mathbf{u} = 0$$

This looks very much like a continuity equation, but now it's on perturbed quantities.

We can play similar games with the Euler and Poisson equation. We get

$$\frac{d\mathbf{v}}{dt} + (\mathbf{u}\cdot\nabla_r)\mathbf{v} = -\frac{1}{\rho}\nabla_x p - \nabla_x \Phi$$

The first term becomes $d\mathbf{v}/dt = d(H\mathbf{x})/dt + (dR/dt)\mathbf{u} + R(d\mathbf{u}/dt)$, and the first term of this cancels the homogeneous part of the potential. The second term here generates two terms, the first of which is $(\mathbf{u}\cdot\nabla_r)H\mathbf{x} = R\mathbf{u}H$. This will have important consequences.

$$\frac{d\mathbf{u}}{dt} + 2H\mathbf{u} + (\mathbf{u}\cdot\nabla_r)\mathbf{u} = -\frac{c_s^2}{R^2}\frac{\nabla_r\delta}{1+\delta} - \frac{1}{R^2}\nabla_r\phi$$

Here we consider the universe is expanding adiabatically, and the perturbation in energy and density are related to the adiabatic sound speed:

$$\partial p/\partial\rho = c_s^2$$

, and Poisson equation becomes:

$$\nabla_r^2\phi = 4\pi G R^2 \rho_h \delta$$

Here $\Phi = \phi_0 + \phi$

These are the full equations for gravitational perturbations in comoving coordinates. They are non-linear.

Now let's consider small perturbations. We keep only linear terms.

$$\frac{d\delta}{dt} = -\nabla_r\cdot\mathbf{u}$$

$$\frac{d\mathbf{u}}{dt} + 2H\mathbf{u} = -\frac{c_s^2}{R^2}\nabla_r\delta - \frac{1}{R^2}\nabla_r\phi$$

$$\nabla_r^2\phi = 4\pi G R^2 \rho_h \delta$$

We take the negative divergence of the Euler equation

$$-\frac{d\nabla_r\cdot\mathbf{u}}{dt} - 2H\nabla_r\cdot\mathbf{u} = \frac{c_s^2}{R^2}\nabla_r^2\delta + \frac{1}{R^2}\nabla_r^2\phi$$

$$\frac{d^2\delta}{dt^2} + 2H\frac{d\delta}{dt} = \frac{c_s^2}{R^2}\nabla_r^2\delta + 4\pi G\rho_h\delta$$

Because this is a linear equation, we expand in spatial Fourier modes: $\delta = D(t)\exp(-i\mathbf{k}\cdot\mathbf{r})$ (obviously, we're assuming that only the real part matters). This gives

$$\ddot{D} + 2H\dot{D} = \left(4\pi G\rho_h - \frac{c_s^2 k^2}{R^2}\right) D$$

$D(t)$ is called the growth function. Note that only ratios of growth functions matter!

3 Jeans' Instability

Before we proceed with cosmology, let's step back and look at the original problem that started these discussion which has many more astrophysical applications. The Jeans' instability posted by Sir James Jeans in 1912, which has been the foundation of astrophysical collapse in many environment and you might have seen a much simpler version of our derivation when studying star formation and galaxy formation before.

The only difference is that now we are assuming a static universe, so H=0, or $dR/dt = 0$. Let's consider a wave in the form of

$$\delta(t) = \delta_0 \exp[-i(\mathbf{k}\cdot\mathbf{r} - \omega t]$$

Solving the second order differential equation, you will have the dispersion relation,

$$\omega^2 = c_s^2 k^2 - 4\pi G\rho.$$

This dispersion relation describes oscillations (stable) or instabilities depending on the sign of its right hand side. Introducing Jeans wavelength

$$\lambda_J = 2\pi/k_J = c_s(\pi/G\rho)^{1/2}$$

- If $c_s^2 k^2 > 4\pi G\rho$, the RHS is positive and the perturbation is oscillatory. They are sound waves in which the pressure gradient is sufficient to provide support. So for perturbation with wavelength shorter than λ_J, they will not grow.
- If $c_s^2 k^2 < 4\pi G\rho$, the RHS is negative, this is the unstable mode, in which the perturbation will either grow or fall exponentially

$$\delta \propto \exp(\Gamma t),$$

Where $\Gamma = \pm[4\pi\rho G(1-\lambda_J^2/\lambda^2)]^{-1/2}$. The positive sign corresponds to growing mode. For very long wavelength, the growth rate $\Gamma \sim (G\rho)^{1/2}$, or the growth time $\tau \sim (G\rho)^{-1/2}$. This is your dynamical timescale for a spherical collapse.

The physics for Jeans instability is simple. We have hydrostatic pressure support:

$$dp/dr = -GM(< r)/r^2,$$

The region becomes unstable when gravity overwhelms pressure support. So $dp/dr \sim p/r \sim c_s^2\rho/r$ and $M \sim \rho r^3$, then we will have for $r > r_J \sim c_s/(G\rho)^{1/2}$, it becomes Jeans unstable. The other way to put it is sound crossing time r/c_s is less than dynamical time $1/(G/\rho)^{1/2}$, then it becomes unstable. We will see Jeans instability again when discussing galaxy formation.

4 Jeans' instability in an expanding universe

Before, we had an equation with time-independent coefficients, so we could use time exponentials. This led to a dispersion relation that indicated that $k < k_J$ grew exponentially.

Now, H and ρ_h both depend on time, so exponentials aren't the correct solution. Since we are worrying out large scale structure, we will only consider long wavelength $\lambda >> \lambda_J$, in which case we neglect the pressure term.

$$\ddot{D} + 2H\dot{D} = 4\pi G\rho_h D$$

Let's consider a matter-dominated universe ($\Omega = 1$) in which $c_s = 0$. This means that $H = \frac{2}{3t}$ and $8\pi G\rho_h = 3H^2 = \frac{4}{3t^2}$. Hence, we have

$$\ddot{D} + \frac{4}{3t}\dot{D} - \frac{2}{3t^2}D = 0$$

This has a form that can be solved by power-laws, so let's try $D = t^n$. This gives

$$n(n-1) + \frac{4n}{3} - \frac{2}{3} = 0$$

which has solutions $2/3$ and -1. This means that we have one mode that grows in time, and another that decays.

$$D = At^{2/3} + Bt^{-1}$$

Next, we remember that $R = t^{2/3}$. So our growing mode has its amplitude growing as R, while the decaying mode decreases as $R^{-3/2}$.

We no longer have exponential growth! The expansion of the universe slows the growth to a power-law! In particular, since the perturbation growth with the exact same dependent on t as R, this means perturbation growth self-similarly as the expansion of the universe, $\delta \propto (1+z)^{-1}$, so in a critical universe, new structure is always forming because the universe is self-similar.

We can also work out the case in which the universe is empty, $\dot{R} = constant$, and $\rho = 0$, i.e., the Milne model, in this case,

$$\ddot{D} + 2H\dot{D} = 0,$$

and $H = 1/t$. Seeking power law solution, we find $n = 0$ or $n = -1$, so there is a constant mode and a decay mode. Perturbation won't grow in an empty universe.

$$D = At^0 + Bt^{-1}$$

In general, one can write down the solution for the growing mode as

$$\Delta(R) = 5/2\Omega_{m0}H \int_0^R dR'/(dR'/dt)^3$$

where R is the scale factor. And for a low density universe, it goes through two stages, at $z >> 1$, the density parameter $\Omega_m \sim 1$, so the perturbation grows linearly. At low redshift, $\Omega_m \sim 0$, so the grow freezes out, with the transition happens at $\Omega_0 z \sim 1$. This gives us yet another way to probe cosmological parameters: by looking at the density of objects in the universe, or the abundance of structures in the universe, if the universe is of critical density, then the number density will decline rapidly towards high-z; if it is low density, then the number density will be roughly constant to $\Omega_0 z \sim 1$ and then decline towards high redshift. The dN/dz test is a very powerful test since the growth factor depends both on Ω_m and Λ through the evolution of scale factor. It is another important way to probe dark energy. It is a particularly powerful probe to use number count of collapsed objects, such as cluster of galaxies, for cosmological test, which we will see later.

5 Peculiar Velocity

The development of velocity perturbations in the expanding universe can be derived from the Euler equation we derived before:

$$\frac{d\mathbf{u}}{dt} + 2H\mathbf{u} = -\frac{c_s^2}{R^2}\nabla_r\delta - \frac{1}{R^2}\nabla_r\phi$$

and ignore pressure term.

$$\frac{d\mathbf{u}}{dt} + 2H\mathbf{u} = -\frac{1}{R^2}\nabla_r\phi$$

Note that this is the comoving peculiar velocity. Let's split the velocity vector into components parallel and perpendicular to the gravitational potential gradient.

$$\mathbf{u} = \mathbf{u}_{\parallel} + \mathbf{u}_{\perp}.$$

The parallel term is referred as potential motion, and the perpendicular part the vortex or rotational motions.

For rotational motions,

$$\frac{d\mathbf{u}_{\perp}}{dt} + 2H\mathbf{u}_{\perp} = 0$$

For $\Omega = 1$, such modes scale in time as $\mathbf{u}_{\perp} \propto R^{-2}$. Hence, they decay. This is important because it means that primordial vorticity is erased. We expect the velocity field today to be curl-free. This poses issue with galaxy formation model involving primordial turbulence.

Neglect perpendicular part, the velocities that match the density modes are proportional to the gradient of the potential. Let's try

$$\mathbf{u} = -F(t)\frac{\nabla_r \phi}{R} = F(t)\mathbf{g},$$

where $\mathbf{g} = -\nabla_r \phi / R$ is the gravitational force. The continuity equation becomes

$$\frac{d\delta}{dt} = -\nabla_r \cdot \mathbf{u} = F(t)\nabla_r^2 \phi / R = F(t) 4\pi G R \rho_h \delta$$

We put in $8\pi G\rho = 3H^2\Omega$, and use $\delta \propto D(t)$. This means that

$$F(t) = \frac{\dot{D}}{D}\frac{1}{4\pi G \rho_h R} = \frac{\dot{D}}{D}\frac{2}{3H^2\Omega R} = \frac{2}{3H\Omega}\frac{dD/dR}{D}$$

$$\delta v = RF\mathbf{g} = \frac{2f}{3H\Omega}\mathbf{g}$$

where $f = (R/D)(dD/dR) \approx \Omega^{0.6}$ for the growing mode.

The growing mode has a velocity that is proportional to the gravitational acceleration. For $\Omega = 1$, $\delta\mathbf{v} = t\mathbf{g}$, it's simply the present-day gravitational acceleration times the age of the universe!

The growing mode of velocity field will map the growing mode of density field. Consider a plane wave density perturbation. $\delta = D(t)\delta_0 \exp(-i\mathbf{k}\cdot\mathbf{r})$. The potential must have the same exponential dependence: $\phi = \phi_0 \exp(-i\mathbf{k}\cdot\mathbf{r})$.

Then we have $-k^2\phi_0 = 4\pi G R^2 \rho_h \delta_0 D(t)$ from the Poisson equation, and

$$\mathbf{g} = -\nabla\phi/R = -i\mathbf{k}(\phi_0/R)\exp(-i\mathbf{k}\cdot\mathbf{r}) = \frac{i\mathbf{k}}{k^2} 4\pi G R \rho_h \delta_0 D(t) \exp(-i\mathbf{k}\cdot\mathbf{r}) = \frac{i\mathbf{k}}{k^2}\frac{3H^2\Omega R}{2}\delta$$

$$\delta \mathbf{v} = \frac{i\mathbf{k}}{k^2} H R f \delta$$

What this means is that the mapping of density and velocity field depends on this factor f, which is related to density parameter mainly. So by mapping the large scale density and velocity field, this gives us yet another way of measuring Ω. Measuring peculiar velocity is not easy. But this is fairly popular in the 1990s, and for a while had produced the only observational evidence that $\Omega \sim 1$. Later, with better peculiar velocity measurements, this seems to have gone away.

6 Non-linear perturbations

So far we've described the evolution of small perturbations. However, this is certainly not the whole story.

When overdensities become non-linear ($\delta \approx 1$), the linear approximation does not break gracefully. Gravity is an attractive force, and the overdensities quickly run away to very large density!

Higher-order perturbation theory is generally a very poor approximation in cosmology. We quickly reach a regime of $\delta \gg 1$.

Numerical simulations show that the matter accumulates in dense regions. We call these regions halos, because of our expectation that they correspond to the dark matter halos of galaxies and clusters.

What stops the density from going to infinity? Random motions of the particles. As the particles enter the dense region from various directions, they interpenetrate (the gas shocks). This produces an effective pressure term that opposes further collapse.

Simulations and analytic models suggest that collapse is halted at an average overdensity of about 200. The interior of the halos have higher densities, but it's not because of cosmological infall.

An important aspect of gravitational collapse is that non-linear collapse on small scales does not spoil the linear evolution of large-scale perturbations. These perturbations don't care whether the matter on small scales is lumpy or not. Credit Gauss's law!

Hence, in the non-linear regime, halos become our standard description.

Halos have different sizes and masses. They can merge together. There are models and simulation results (that agree) to describe the mass functions and the merger rates.

Halos are viewed as the sites of galaxy and cluster formation. Hence, the formation of galaxy-mass-sized halo is the precursor to the formation of luminous galaxies, while the merging of those halos is the precursor to galaxy merging and cluster formation. We'll talk more about this later on.

One cannot solve the gravitational collapse problem from arbitrary initial conditions in the non-linear regime. There are some analytic models (some of which have been remarkably useful), but the primary workhorse here are numerical simulations.

In the simplest form, one represents the cosmological density field by millions of particles and then follows the motions of these particles according to their self-gravity. At the initial time, the particles are nearly on a grid, but have been either displaced in position or in mass to install some small perturbations, randomly generated to mimic the linear regime power spectrum of the chosen cosmological model.

The particles generically clump into halos and one can study the inner properties and the external correlations of these halos.

However, comparing to the real world requires building some statistics, so this is what we'll talk about next!

- $\delta\rho/\rho = \Delta \ll 1$: **linear perturbation.**
- $\delta\rho/\rho = \Delta \gg 1$: **non-linear, simulations.**
- $\delta\rho/\rho = \Delta \sim 200$: **collapsed, viralized objects, "halo".**
- **halo formation → galaxy formation**
- **halo merger → galaxy merger and cluster formation**

7 Correlation Function

From these redshift surveys we know that galaxies are clustered. So how do we describe the clustering mathematically. Here I am going to introduce two very important statistics. Correlation function, and power spectrum.

One way to describe the tendency that galaxies like to stay together is the two point correlation function $\xi(r)$. Correlation functions $\xi(r)$ is the excess probability that a galaxy is found at a distance r from a known one. Hence,

$$\xi(r) = \left\langle \frac{\rho(\mathbf{x}+\mathbf{r})\rho(\mathbf{x})}{\rho_0^2} \right\rangle - 1 = \langle \delta(\mathbf{x}+\mathbf{r})\delta(\mathbf{x}) \rangle$$

Here $\delta(r)$ is the density fluctuation field of galaxies in the universe

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \rho_0}{\rho_0}.$$

It is the density contrast to the random density. You might have thought the argument should be a vector, but the claim is that the universe is statistically isotropic.

This idea of a correlation function can be extended to triplets and so forth. These are called higher-order correlation functions. Note that these depend on the geometry of the locations, so the vector directions of the separations now matter.

The other way to put it: let us have two small volumes ΔV_1 and ΔV_2, separated by distance of r. The average spatial density of galaxies is ρ. Then the number of galaxies in ΔV_1 is $\rho\Delta V_1$. Since ρ and ΔV are small numbers, it means that the chance of finding a galaxy in ΔV_1 is $\Delta P_1 = \rho\Delta V_1$. Now if galaxies are truly randomly distributed, then the chance of find a galaxy in ΔV_2 is $\Delta P_2 = \rho\Delta V_2$. So the chance of finding a galaxy in both ΔV_1 and ΔV_2 is

$$\Delta P = \rho^2 \Delta V_1 \Delta V_2.$$

In this case, galaxy distribution is uncorrelated. If galaxies are clustered, or correlated, then the chance of finding a galaxy in ΔV_1 and at the same time in ΔV_2 is bigger than random:

$$\Delta P = \rho^2 \Delta V_1 \Delta V_2 (1 + \xi(r)).$$

So $\xi(r)$ describes how much the two volumes, separated by r, are correlated, or how strong the galaxies are clustered. If $\xi(r) > 0$, they are clustered. If $\xi(r) < 0$, they are anti-clustered, in other words, they tend to avoid each other. Clearly, in order to compute $\xi(r)$, we need to measure redshift in order to derive the real space distance. Also clear is that ξ is a function of distance separation. At very large distance separation, the galaxy distribution will have nothing to do with each other, so $\xi(r) \to 0$ when $r \to \infty$. On the other hand, if galaxies are clustered, then the closer the separation is, the higher the possibility of finding a galaxy next to a known galaxy is, comparing to the random field. When $r \to 0$, this probability is going to go to 1, or $\xi(r) \to \infty$. From observations, people find that it is a very good power law:

$$\xi(r) = (r/r_0)^{-\gamma}.$$

Here r_0 is called the correlation length, the scale at which the correlation function is 1, or the probability of finding a galaxy from a known galaxy at a distance r is twice as high as finding a galaxy in the random field. It is a very important parameter to measure in galaxy surveys. At a few times r_0, the galaxy distribution becomes to be very close to random.

Most galaxy redshift survey shows that $r_0 \sim 5/hMpc$ and $\gamma \sim 1.8$.

Deviation from this power law on both small and large scales.

- small scale: non-linear growth of halo will matter. Two halo term. ($< 1Mpc$).
- large scale: can't ignore that the universe is not pure dark matter. Baryon in the early universe causes acoustic waves. BAO. $> 100Mpc$.

8 Power Spectrum

A common reformulation of these ideas is to study the density field in Fourier space. First, we Fourier transform

$$\hat{\delta}_{\mathbf{k}} = \int d^3x \, \delta(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}}$$

and then we consider the statistical properties of the Fourier coefficients (which are complex numbers).

The translational invariance of the statistical distribution gives useful properties. In particular, $\left\langle \hat{\delta}_{\mathbf{k}} \right\rangle = 0$.

The most common statistic in Fourier space is the power spectrum. This is the two-point correlations of the Fourier coefficients. In particular, we have

$$\left\langle \hat{\delta}_{\mathbf{k}} \hat{\delta}^*_{\mathbf{k}'} \right\rangle = (2\pi)^3 \delta^3_D(\mathbf{k}-\mathbf{k}') P(k)$$

$\delta^3_D(\mathbf{k}-\mathbf{k}')$is the Dirac delta function.

The power spectrum and correlation function are really just two representations of the same information. In fact, the two functions are 3-d Fourier transforms of one another.

$$\delta(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \delta_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}}$$

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{r})\delta(0) \rangle = \langle \delta(\mathbf{r})\delta(0)^* \rangle = \left\langle \int \frac{d^3k}{(2\pi)^3} \delta_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}} \int \frac{d^3k'}{(2\pi)^3} \delta^*_{\mathbf{k}'} \right\rangle$$

$$= \int \frac{d^3k}{(2\pi)^3} \int \frac{d^3k'}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{x}} \langle \delta_{\mathbf{k}} \delta^*_{\mathbf{k}'} \rangle = \int \frac{d^3k}{(2\pi)^3} \int \frac{d^3k'}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{x}} (2\pi)^3 \delta^3_D(\mathbf{k}-\mathbf{k}') P(k) = \int \frac{d^3k}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{x}} P(k)$$

While power spectrum and correlation function are basically the same thing, they are measured in quite different ways, one by counting galaxy pairs, the other by deriving the density field and its Fourier transform. They have different error properties and can have pros and cons each in real life measurements.

What is a power spectrum, really? It's the mean square amplitude of perturbations of a given wavenumber. Since we found that each Fourier mode evolves independently in linear theory, power spectra are particularly useful on linear scales.

A useful quantity to compute is the rms density fluctuation in a sphere of radius R. To do this, first write that the density fluctuation in the sphere is

$$\Delta_R = \frac{3}{4\pi R^3}\int_{sphere} d^3r\ \delta(\mathbf{r})$$

This has

$$\langle\Delta_R\rangle = \frac{3}{4\pi R^3}\int_{sphere} d^3r\ \langle\delta(\mathbf{r})\rangle = 0$$

The rms is then $\sigma(R) = \langle\Delta_R^2\rangle^{1/2}$.

For notational simplicity, let $W(\mathbf{r}) = 1$ if the point is inside the sphere and $W(\mathbf{r}) = 0$ otherwise. W is called the window function. This special case is called a top-hat window. Then

$$\sigma^2(R) = \left\langle\Delta_R^2\right\rangle = \left\langle\frac{1}{V}\int d^3r\ \delta(\mathbf{r})W(\mathbf{r})\frac{1}{V}\int d^3r'\ \delta(\mathbf{r}')^*W(\mathbf{r}')\right\rangle = \frac{1}{V^2}\int d^3r\ d^3r'\ W(\mathbf{r})W(\mathbf{r}')\xi(|\mathbf{r}-\mathbf{r}'|))$$

Without going through a few lines of simple but long math that deal with changing integrals in Fourier transformation, we will get a simple one dimensional integral

$$\sigma^2(R) = \int \frac{d^3k}{(2\pi)^3}P(k)\frac{|W_\mathbf{k}|^2}{V^2}$$

where $W_\mathbf{k}$ is the Fourier transform of our window function.

Today, the rms fluctuation in spheres of $8h^{-1}$ Mpc radius is roughly 1. This corresponds to masses of about $2\times10^{14}h^{-1}$ $\mathrm{M}_\odot$, which not coincidentally is the typical cluster mass. This is yet another fundamental cosmological parameter: σ_8.

Note that transforming between power spectrum, which tells you the amplitude of fluctuations in different Fourier modes, and measured density fluctuation, involves details of your window function. In observations, this window function is the geometry of your survey, including both angular shape and vertical completeness as a function of redshift. Deriving power spectrum, which is what we really care as a physical quantity to be compared with theoretical predications from the real observations requires one to understand the survey window function and model it carefully. And ideally, you want to design your survey to have as simple window function as possible. But it is almost never possible to have a volume limited survey with a top-hat window function.

It is important to note that the contribution to σ goes as $\int \frac{dk}{k} k^3 P(k)$. This means that $P(k)$ is *not* the appropriate quantity to base one's intuition on where fluctuation power is coming from. Rather, it is useful to think in terms of logarithmic intervals, in which case $k^3 P(k)$ is the relevant quantity. Basically, P is the fluctuation per point in k space, but the number of points in k that contribute to the fluctuations scales as k^3.

Indeed, it is increasingly popular to define $\Delta^2(k) = k^3 P(k)/2\pi^2$ as the power spectrum (I call it the dimensionless power spectrum). We then have $\sigma^2 = \int \frac{dk}{k} \Delta^2(k) |W_{\mathbf{k}}|^2$.

Window functions are defined to have $W_{k=0} = 1$, and they then usually cut off at some particular k. For most power spectra of interest, Δ^2 is increasing with k, and so it is a rough approximation that $\sigma^2 \approx \Delta^2$ where the latter is evaluated at the the cutoff scale of W, or $\sigma^2 \approx \Delta^2(1/R)$.

But one important issue in cosmology is that galaxies may or may not trace mass because of the presence of dark matter. And in cosmology we mostly care about matter. So the real parameter we concern about, and the parameter that theory is going to tell us, the the mass fluctuation $\sigma_8(mass)$. It may or may not be one. And the ratio of the fluctuation in the counts of galaxies to the fluctuation of underlying mass distribution is called the bias factor:

$$b = \frac{\sigma_8(\text{galaxies})}{\sigma_8(\text{mass})}.$$

Since we know $\sigma_8(gal) = 1$, to determine bias factor is the same as to determine $\sigma_8(\text{mass})$, or to determine the normalization of the mass power spectrum. A lot of effort in galaxy survey business is to determine both the shape and the normalization of galaxy power spectrum, and compare them with simulations, to constrain both cosmological parameters, as well as different flavors of dark matter models, in particular the thermal properties of dark matter particles.

9 Types of dark matter

We know that most matter in the universe is dark matter. Let us first recall the reasons for talking about dark matter, in particular non-baryonic dark matter, seriously in the context of galaxy formation.

There are mostly three lines of evidence:

(1) the mass to light ratio in the MW, in galaxies, and in cluster of galaxies is much to high for normal stellar population, in other words, the dynamical mass of a stellar system bigger

than globular cluster is always much bigger than stellar mass. So there must be missing, or dark, matter.

(2) the determination of cosmological parameters should that $\Omega_m \sim 0.3$, and from BBN, $\Omega_b < 0.05$, so there must be non-baryonic dark matter.

(3) a purely baryonic model of galaxy formation could not fit CMB data, nor could it generate the structure that we see today.

So most people believe that the mass of the universe is dominated by non-baryonic dark matter, i.e. mass in forms other than quarks, which makes protons and neutrons.

There are two flavors of DM people are still considering:

1. Hot dark matter. These are particles that decouple from the rest of the universe when it was relativistic, and which have a number density roughly equal to that of photons. In this case, there rest energy is a few eV, particular, few ev-mass neutrinos. Why we call them "hot"? For neutrinos with T = 1.95K today (neutrino background), their velocity is going to be: $v = 158(m/eV)^{-1}$km s^{-1}. So the thermal velocity is very high, and it is even higher at high redshift, this leads major effects on the development of self-gravitating structures.

2. Cold dark matter. It the particles decouple while they are nonrelativistic, there thermal velocity today is effectively zero, and thus "cold". The typical candidates are WIMPS, or weak-interacting massive particles, with possible mass of a few GeV.

One of the key considerations for theories of structure formation in which non-baryonic dark matter dominant, is the damping of density perturbation by free-streaming. So long as the DM particles are strongly coupled, they behave no different from ordinary particles. At later epochs, however, DM particles no longer interact with other particles. If the particle were relativistic at the froze-out tome, they would continue to travel in straight lines at the speed of light. Thus, if the particles belonged to some density perturbation, as they continue to travel freely, they are going to damped out, or reduce the density fluctutation. The masses which are damped out depend upon how far the free-streaming particles can traevel at a given epoch. The comoving distance which a free-streaming particle can travel by epoch t is just:

$$r_{FS} = \int_0^t v(t)/R(t)dt.$$

Note that before the decoupling, the particle has the speed of light; however, after the decoupling the particle will start to cool down as its speed will slow down with $1/(1+z)$. Taking this into account, one find that the mass scale of free-streaming damping is:

$$M_{FS} = 4 \times 10^{15}(m/30eV)^{-2}M_\odot.$$

Therefore, if the hot dark matter neutrino mass is 30 eV, all density perturbations on mass scale less than this will be damped out as soon as they come through the horizon. Since these masses are of the order of massive clusters, in this picture, then, only structures at larger scales can survive. There will be no or little power in the mass power spectrum at smaller scales. So in hot dark matter cosmology, you have to make very large structure, the so-called "zeldovich pancake" first, and then the pancake will break down to make small structure. This scenario is called the "top-down" structure formation.

On the other hand, in the cold dark matter scenario, the CDM particles decoupled early in the universe, after they had become non-relativistic. Free-streaming is unimportant. So in CDM model, all the initial perturbations are preserved. The CDM model is the popular, and working model of structure formation.

The CDM model can be studied using computer simulations in great detail, once the initial power spectrum of fluctuation is given. As I will show you in a moment, these spectra are such that there is most power on the small scale, so the lowest mass objects form first. These then undergo hierarchical clustering under the influence of perturbations on large scales and so the larger scale, which has smaller initial perturbation, and therefore takes longer time to grow, are built-up later. This bottom-up, or hierarchical structure formation is the base of all the galaxy formation theory today.

10 Primordial power spectrum

The density fluctuation power spectrum of the universe consists of two parts: the primordial power spectrum, which is power spectrum the universe began with, and the effect that happens on top of that after the scale in question has entered the horizon as the universe is expanding. The function that describes the changes to the input initial spectrum of the perturbation is called the transfer function.

The initial power spectrum is typically described as having a power-law form:

$$P(k) \sim k^n.$$

Let's call the density fluctuation at mass scale M to be $\sigma(M)$, one can show (find proof in Longair) that: $\sigma(M) \sim M^{-(n+3)/6}$, so as long as $n > -3$, the mass fluctuation decreases to large mass scales. This is good. We want the universe to be uniform on large scales.

The $n = 1$ case is of special interests. This spectrum has the property that the density fluctuation $\sigma(M)$ had the same amplitude on all scales when the perturbations came through the horizon.

It is useful to look at the density perturbation of a certain size/mass scale at the time it enters the particle horizon $r = r_H = ct$ where t is the age of the universe at a certain time. We mentioned before that in radiation dominated era, the perturbation grows as $\delta \sim R^2$. So the fluctuation grows and changes with horizon mass scale as:

$$\sigma(M_H) \sim R^2 M_H^{-(n+3)/6}.$$

And the horizon mass scale is $M_H \sim \rho_D t^3$. In radiation dominated era: $R \sim t^{1/2}$, $\rho \sim R^{-3}$, so $M_H \sim R^3$. Plug all these things in, you will find:

$$\sigma(M_H) \sim M^{2/3} M^{-(n+3)/6.}$$

And if $n = 1$, the amplitudes of the density perturbation were all the same when the came through their particle horizons during the radiation dominated era.

This particular form of primordial power spectrum is called the Harrison-Zeldovich spectrum, first suggested by Harrison and Zeldovich, independently, in early 1970s. It has a number of very appealing features. In particular, if $n = 1$, then the universe is fractal, or self-similar, in the sense that every perturbation came through the horizon with the same amplitude, and as the universe expands, we always find perturbations of the same amplitude appearing on the horizon; the universe looks the same when viewed on the scale of the horizon. It is simple, and it occurs very naturally in inflationary models. It is sometimes also called the scale-invariant spectrum. Deviations of the power-law exponent from this special value are known as 'tilts'. Modern measurements find that n is rather close to 1, with a measurable tile, e.g., Planck finds that $n = 0.97 \pm 0.01$.

11 Transfer Function

Transfer function, or the effect on the perturbation power spectrum after it has entered horizon. The transfer function T(k) describes how the shape of the initial power spectrum in the dark matter is modified by different physical processes:

$$\Delta_k(z = 0) = T(z) D(z) \Delta_k(z)$$

Here D(z) is the growth function we discussed before $D(t) \propto 1/(1 + z)$ for critical universe. There are two effects that I want to mention:

(1) the free-streaming effect that we just talked about. This will reduce the small scale power in hot DM models because the gravitationally important neutrinos are free-streaming around and damped the fluctuation.

(2) effect related to pressure in radiation dominated era. In the matter-dominated era, all scales grow equally. However, in the radiation-dominated era, things behave differently. Here, pressure is important; indeed, the sound speed is $c/\sqrt{3}$! Remember that in the previous derivation of growth function, we ignored the sound speed and pressure term by assuming CDM. But it does matter in early times due to coupling. On scales smaller than the horizon, the growth is stalled by the presence of the radiation pressure. Basically, the universe expands too quickly for the dark matter to collapse. So on small scales the radiation remains smooth, but on large scales it has to cluster. So, scales that enter the horizon (i.e. suffer from smooth radiation) during the radiation-dominated epoch grow less than those that enter the horizon during the matter-dominated epoch. As soon as the perturbations came through the horizon, they cease to grow until the epoch of equality, after which they grow as described in the growth function.

One thing we didn't do is to work out the growth of perturbation in radiation dominated era. Going through the same process as we did before, one can show that in radiation dominated era,

$$\Delta \propto t \propto (1+z)^{-2}.$$

Thus, for the small scale (large k) which enters horizon early, Δ will not grow and suppressed by a factor of $(R_H/R_{eq})^2$, or by a factor of k^2. Therefore, the power spectrum itself will be suppressed by a factor of k^4 on small scale or large k. The transition happens at the horizon size of matter/radiation equality, which is about 100 Mpc co-moving.

This effect results in that the density fluctuation $\sigma(M) \sim M^{-(n+3)/6}$ is roughly flat on small scales. So the processed power spectrum will have $P(k) \sim k^{-3}$ on small scales, $P \propto k$ on large scales, the scale of the horizon at matter-radiation equality, which is about 100 Mpc for LCDM model in comoving units.

12 Power spectra from redshift survey

SDSS Figure.

The measurement of power spectra from galaxies and other cosmic structures have provided a powerful probe to cosmology.

(1) the overall shape of power spectrum provide information on the initial power spectrum, whether it is Harrison-Zeldovich, or tilted.

(2) The small scale shape shows that the perturbation is mostly from CDM. HDM has very

different shapes.

(3) The peak of power spectrum corresponds to the horizon scale of equality of mater and radiation. This wave number is related to both the redshift of equality, and the size of the universe at that time which is related to the expansion history. Turns out $t_{eq} = 7.3 \times 10^{-2} \Omega_0 h^2$ Mpc^{-1}. Therefore, measurement of power spectrum directly measures $\Omega_0 h$ – the other factor of h is absorbed in the distance dependence of Hubble constant.

(4) There are detailed difference between open and Λ dominated models, but they are subtle. The power to determine Λ from LSS comes from combining with CMB which shows that the universe is flat.

AST541 Lecture Notes: CMB March 2024

The power spectrum of angular fluctuations in the CMB and its polarization properties provide a wealth of information of cosmological significance, both for the determination of cosmological parameters and for understanding the structure formation in the Universe. Starting with ground-based and balloon experiments, the experimental success was highlighted by COBE, WMAP and Planck missions, Nobel prize worthy experiments.

Our goal here us to related the models of structure formation to the imprint they leave on the CMB. COBE, WMAP and Planck, combined with low-z redshift survey, bring us to the era of precision cosmology, in which we are now able to determine cosmological parameters to better than 10%.

We will not discuss in great details about the experimental aspects of things. It is a highly specialized field. The experimental achievements now is at the precision that requires very detailed modeling. To include all the relevant physics, the only feasible approach is to find numerical solutions of the coupled Einstein - gravity, Boltzmann - statistical, and fluid dynamical equations. There are now great off-the-shelf codes to do this for you, for example CMBFast. If you want to interpret CMB results, you will need to rely on those. Our goal here is to outline the general physical processes, and to explain the general principles of why we can use CMB anisotropy to determine cosmological parameters.

We will first establish the basic scales of our problem. Then introduce the tools CMB people uses to describe the angular anisotropy. Next class, we will talk about features on the CMB power spectrum and results on cosmological parameters.

1 The Recombination Era

As shown before,

$$\tau = 0.035 \Omega_B / \Omega_m^{1/2} h z^{3/2}$$

where τ is the optical depth due to Thompson scattering. So at $z \sim 1000$, optical depth is very large. Any radiation originating from higher redshift than the epoch of recombination was scattered many times until the last scattering surface. This is very similar to the problem of stellar atmosphere. The redshift at which $\tau \sim 1$ corresponds to the stellar photosphere. One can define a visibility function:

$$v(z) = e^{-\tau} d\tau/dz,$$

which tells you at which redshift most of the photons reach us.

Figure.

We discussed a little bit about calculation of recombination before. One needs to include carefully all processes about Lyman line and continuum photons, in particular two-photon process in which two photons are liberated from the 2s state to the ground state in a rare quadrupole transition, which turns out to be the dominant way to get hydrogen to the ground state, because one photon process will generate a high energy photons that can be used immediately to ionize another hydrogen at large optical depth.

The important thing about this figure is that recombination, or last scattering, didn't happen at a single redshift. Rather, for a maximum visibility at z=1090, the half maxima are at 1178 and 983, with an interval of $\delta z = 195$. It takes about 120,000 years. This finite depth of large scattering layer has important consequence.

2 Scales of Fluctuations

Before we look at the CMB power spectrum, let's work out a few physical and angular scales that are important to the CMB first. Here we assume a LCDM cosmology with concordance numbers.

2.1 The last scattering layer

Taking the thickness of the last scattering layer to correspond to a redshift interval of 195 at z=1090, the comoving radial distance is then

$$\Delta r = 16.2(\Omega_m h^2)^{-1/2} = 42\text{Mpc}.$$

THe mass within this cale is roughly $2 \times 10^{15} M_\odot$, which is the mass of a cluster. The angular size is $\theta \sim 3'$.

On comoving scale less than 42Mpc at current epoch, we expect a number of independent fluctuations to be present along the line of sight through the last scattering layer. So the random superposition of these perturbations leads to a statistical reduction in the amplitude of the observed intensity fluctuation by a factor of $N^{-1/2}$.

2.2 The sound horizon and acoustic peaks

We briefly mentioned acoustic peaks in the last lecture. The existence, and the properties of acoustic peaks are the most fundamental predictions of CMB anisotropy. We want to calculate the sound horizon at the last scattering, which is defined to be $\lambda_s = c_s t$, where c_s is the sound speed and t is the age of the universe at that time. This is the maximum distance which sound waves could travel and undergo coherent oscillations, therefore, it is the longest wavelength which acoustic waves could have at the epoch of recombination. The sound speed of a photon fluid is $c/\sqrt{3}$. When there are baryons, it slows down a bit. But roughly, $c_s \sim 0.5c$. The comoving size of sound horizon is given by

$$s = \int_0^{t_*} dt\; c_s(z)\;(1+z) = \frac{c}{\sqrt{3}}\int_0^{t_*} dt\;(1+z)$$

For our standard cosmology, this is about 100 h^{-1} Mpc, and it corresponds to ~ 0.6 deg on the sky. This wavelength corresponds to the first maximum in the power spectrum of temperature fluctuations. It is when a single wavelength of oscillation between the entry of horizon and the epoch of recombination. Because after recombination, the acoustic wave could not grow anymore and will freeze. So we are predicting that there will be a very pronounced peak in the CMB power spectrum at this scale, which is the same thing we saw in the galaxy LSS. There will be higher order resonance at smaller scales. The acoustic scale is a standard ruler, or really a sounding rod. Note that we are only talking about baryons. There will be a background of perturbation from dark matter which is a result of our Harrison-Zeldovich spectrum. The baryon features discussed here are all on top of it.

2.3 Silk damping scale

Another important physical process related to perturbations in baryons is the effect of Silk damping, which is basically the photon diffusion, or dissipation process of the baryon-photon fluid in the pre-recombination era.

Although at that time, matter and radiation were closely coupled, the coupling was not perfect and photons could diffuse out of the density perturbations. Perturbations in baryon density travels as acoustic waves, or sound waves, in the universe. Radiation pressure provides the restoring force which maintained these oscillations. If the photons diffused out of the perturbations, the wave will be damped out. These was first recognized by Joe Silk in 1968. It was only observed in the CMB small scale power spectrum very recently. We want to find out the scale to which photons can diffuse. At smaller scales, baryon perturbation will be damped. The radial distance over which photons can diffuse is

$$r_S \sim (Dt)^{1/2} = (1/3\lambda ct)^{1/2},$$

where D is the diffusion coefficient which goes roughly as the product of mean free path of photons due to Thomspon scattering, the reason that photons will diffuse, and the sound speed of the relativistic fluid. Plug in the age of the universe at recombination, we find that Silk damping scale is 9 Mpc, and the angular scale is about $\sim 5'$ on the sky. What Silk damping is going to do is to damp out high order acoustic peaks in the CMB. There is a homework problem asking you to work out some more details.

2.4 Particle Horizon Scale

Finally, just to remind you that one can calculate the scale of particle horizon at recombination. This is the maximum distance in the universe at which causal connection can be established. We worked out these scale before, it is just a few times larger than the sound horizon. $r_H \sim 500Mpc$ in comoving units, and angular scale of about 2 deg on the sky.

So to ramp up:

- recombination time
- particle horizon: scale large than 2 deg not conected.
- scale between 0.01 and 2 deg, lots of baryonic features, affected by silk damping and finite thickness of last scattering layer. Important cosmological probe,

3 Power Spectra of the CMB

Now we introduce the technique and nomenclature to express the fluctuation in the temperature of the CMB. We are observing a temperature distribution over the surface of a sphere, and we are interested in very large angles. So we need the two-D polar equivalent of the relation between density distribution and power spectrum. The standard technique is to use the spherical harmonic functions as our orthonormal base functions. We first make a spherical harmonic expansion of the temperature distribution of the whole sky:

$$\frac{\Delta T}{T}(\theta, \phi) = \frac{T(\theta, \phi) - T_0}{T_0} = \sum_{l=0}^{\infty} \sum_{m=-1}^{m=l} a_{lm} Y_{lm}(\theta, \phi),$$

where the normalized function Y_{lm} is given by:

$$Y_{lm}(\theta\phi) = \left[\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(sl+|m|)!}\right]^{1/2} P_{lm}(\cos\theta) e^{im\phi} \times \frac{(-1)^m, m \geq 0}{1, m < 0},$$

here Y_{lm} is called a spherical harmonic function of degree l and order m, and $P_{lm}(\cos\theta)$ is an associated Legendre function. Note that for each l, there are 2l+1 values of m. The power of each mode (l,m) is a_{lm}, which can be found as:

$$a_{lm} = \int \frac{\Delta T}{T}(\theta, \phi) Y^*_{lm} d\Omega$$

where Y^*_{lm} is the complex conjugate of the spherical harmonic function. Now we have the two-D power spectrum. Roughly speaking, the degree l is related to the angular separation as $\theta \sim \pi/l$. Therefore l is oftern referred to as the multipole moment, $l = 1$ is the dipole, $l = 2$ is the quadrupole etc.

a_{lm} is a complete description of the density fluctuation. But how is that related to our 3-D power spectrum? In general, the temperature distribution over the sky need not to be Gaussian. If it were Gaussian, then the distribution of a_{lm} for the same l will have random phases, Inflation strongly predict a random phase distribution, since there the origin of fluctuation is quantum fluctuation in the early universe. Other theories that invoking seeding of fluctuations by topological defects, cosmic strings or cosmic textures, suggest non-Gaussian distribution, as they cause abrupt temperature discontinuities. In this case, a_{lm} will be correlated.

If it were Gaussian then each a_{lm} represents an independent estimate of the power at multipole l, so we just average them, and introduce the power spectrum C_l:

$$C_l =< |a_{lm}|^2 >$$

C_l will be a complete description of the temperature fluctuation. It can be shown that for a Harrison-Zeldovich power spectrum,

$$C_l \propto 1/l(l + 1)$$

so you usually see people plot $l(l + 1)C_l$, because on large scale, it will be relatively flat.

one important issue this discussion brings up is the issue of **cosmic variance,**, which ultimately limits the precision of our estimates of temperature fluctuations. The spherical harmonic analysis shows the we obtain (2l+1) independent estimates of the value C_l for a given multipole. The precision with which value of C_l is known is:

$$\sigma(C_l)/C_l \sim [2/2l + 1]^{1/2}.$$

This is the result of we do not have enough sample of the universe, and can not be overcome by better precision in our measurements. So there is a limit that we can know about the property of the largest scale structure in the universe. For COBE, at $l < 20$, the measurements are cosmic variance rather than noise limited, For WMAP, $l < 400$ are cosmic variance limited. In other words, for $l < 20$, COBE is basically as good as it will ever get and future experiment will not improve upon it. This does cause problems. For example, our CMB features a small quadrupole, by 1.5σ or so, comparing to best-fit models. Is there anything wrong with our cosmology, that there is less quadrupole? Or is it just because when you look at many multipoles, some of them by definition will fall out of 1-σ?

Another note: you hear people working on galaxies talking about cosmic variance a lot. Most of the time they use the term wrong. They sometimes say that HDF is too small and the measurement is subject to cosmic variance, since there is strong clustering at that scale, and one is not getting a fair sample of the universe by looking only at a small field. But this is not cosmic variance. The correct term is **sample variance**. Cosmic variance basically says that we have only one universe, and at certain level, our homogeneous cosmology breaks down because we don't have enough independent measurements to figure out what the average value of certain quantity of the universe actually is.

4 CMB power spectrum

Figure 1 Dipole,

Moving through a bath of blackbody photons creates a temperature dipole: hotter in one direction, but does remain a blackbody. The universe does have a preferred frame. Intrinsic dipole would be 10^{-5}, not 10^{-3}. heliocentric dipole has an amplitude of 3.353mK (350km/s), towards $l = 264$, $b = 48$. Part of this is earth rotation in the Galaxy. Correcting for that, LG motion is 600 km/s toward (l,b) = 270, 30; RA=10.5h, dec -26. This is peculiar velocity of the LG.

Figure 2.

Dipole was discovered in early 70's, but other anisotropies weren't discovered until 1992, COBE. At 10 deg scale, CMB fluctuation is about 30μK, or about 10^{-5}.

To get to the picture of WMAP, one has to get rid of: (1) the free-free from galaxy; (2) point sources. One does that by observing in multi-frequency, and then filter out foreground.

Figure.

Finally, CMB power spectrum. Features:

- large scale: relatively flat: a result of Harrison-Zeldovich power spectrum. One can fit the initial power spectrum very precisely, and find out if it is scale invariant, or with a little tile, as inflation predicts. It also somewhat depends on expansion history.
- middle scale: strong acoustic peaks. The first location of the first peak is basically the

angular diameter distance to the large scattering sound horizon. This is determined by a combination of Ωh^2 and expansion history, with small dependence on baryon.

- small scale: small scale effects are complicated, including the silk damping, contribution from clusters of galaxies etc.

So by fitting C_l carefully, one can infer a lot of cosmological parameters. Even more powerful when combining with low-z constraints.

5 Large Scale: Sachs-Wolf effect

On the very largest scales, the dominant source of temperature fluctuation results from the fact that the photons we observe originated from the perturbations at the last scattering layer. Here we consider things bigger than the baryonic acoustic peak and larger than the thickness of the last scattering layer. The only effect one needs to consider, except reionization, is the fact that the photons will pass through gravitational potential fluctuations, and therefore they will be gravitationally redshifted. In general, one can show that as far as the perturbation grow linearly with redshift, what they gain by falling into them is exactly compensated by the redshift coming out. So it is the escape from density perturbations at the epoch of recombination which provides the gravitational redshift. This is known as the Sachs-Wolfe (1967) effect.

A potential well has two effect:

(1) the photon is redshifted as it climbs out of the well

$$\delta T/T = \delta\Phi/c^2$$

(2) the region is time dilated, $\delta t/t = \delta\Phi/c^2$. Clocks in that region run slow relative to the outside world, so that region is younger. This means that it is hotter. Since during matter dominated era, $R \propto t^{2/3}$, $\delta R/R = 2/3\delta t/t$, and $\delta R \sim \delta T$, so the region has $\delta T/T = -(2/3)\delta\Phi/c^2$. So the total is $(1/3)\delta\Phi/c^2$.

So what's the potential perturbation $\delta\Phi$?

$$\Delta\Phi \sim \frac{G\Delta M}{d},$$

and mass fluctuation $\Delta M \sim \Delta\rho d^3$, where $\Delta\rho$ is the amplitude of the density fluctuations, $\Delta\rho = \Delta \times \rho_0$, where Δ is the relative density fluctuation that in $\Omega = 1$ universe, goes like

$\Delta \sim (1+z)$, and $\rho_0 \sim (1+z)^{-3}$, so:

$$\Delta\rho = \Delta\rho_0 (1+z)^2,$$

And the size of the fluctuation $d = d_0/(1+z)$, therefore,

$$\Delta\phi \sim G\rho_0 d_0^2,$$

the perturbation of gravitational potential is independent of the cosmic epoch as long as density perturbation grows as 1/(1+z). This justified our earlier statement that we only need to consider the potential when the photon climbs out of recombination epoch.

We can also use the results from the power spectrum analysis we had before, for $P(k) \sim k^n$: $\Delta \sim M^{-(n+3)/6}$, therefore, $\Delta\rho_0 \sim \rho_0 M^{-(n+3)/6}$, and $M \sim \rho_0 d_0^3$, to derive:

$$\Delta\phi \sim \delta\rho_0 d_0^2 \sim d_0^{(1-n)/2} \sim \theta^{(1-n)/2},$$

so the large scale temperature fluctuation is then:

$$\Delta T/T \sim 1/3 \Delta\phi/c^2 \sim \theta^{(1-n)/2},$$

and for Harrison-Zeldovich, one has a flat temperature power spectrum.

This was detected by the COBE satellite. If one ascribes the power observed to gravitational fluctuations, then one gets a normalization of the power spectrum of fluctuation at Gpc scales at z = 1000. This is already a powerful constraint on models of structure formation. It is a striking aspect of cold dark matter cosmologies that they can reproduce the level of the CMB and the level of large-scale structure today.

It should be noted that the independence of the power spectrum on muliple moments depends on the assumption that the development of the perturbation is linear from the recombination to current epoch. For models with $\Omega \neq 1$, in particular for ΛCDM model, this is no longer true at small redshift. In this case, we need to integrate Sachs-Wolfe effect over all redshift back to recombination. The integrated SachsWolfe effect is also caused by gravitational redshift, however it occurs between the surface of last scattering and the Earth, so it is not part of the primordial CMB. It occurs when the Universe is dominated in its energy density by something other than matter. If the Universe is dominated by matter, then large-scale gravitational potential energy wells and hills do not evolve significantly. If the Universe is dominated by radiation, or by dark energy, though, those potentials do evolve, subtly changing the energy of photons passing through them. A signature of the late-time ISW is a non-zero cross-correlation function between the galaxy density and the temperature of the CMB. It was observed in the SDSS, lending additional evidence to dark energy.

6 Intermediate Scale; the Acoustic Peaks

Plasma before recombination has lots of pressure from the photons. This allows it to oscillate rather than collapse.

$$\delta\phi = \phi_* - \phi_0 = \int_0^{t_*} c_s k(1+z)dt = ks$$

Hence, all modes of a given k have the same phase at recombination. We have fluctuations in the early universe that have a characteristic scale. In the limit that $\Omega_b h^2$ is small and $\Omega_m h^2$ is large, the sound horizon can be computed to be

$$\frac{2c}{H(z_*)}\frac{1}{\sqrt{3}}(1+z_*)$$

We see this scale projected on the sky at the angular diameter distance to $z = 1000$.

It turns out that the angular scale depends very strongly on the curvature of the universe and only weakly on Ω_m and Λ. Open universes would make the scale smaller. Hence, the location of the acoustic peaks in the CMB are a very precise test of the curvature of the universe!

We just discussed the monopole mode of the acoustic peak. Multipoles also present. One can show that, for the temperature fluctuation $\Theta = \Delta T/T$,

$$\frac{d^2\Theta}{dt^2} = -1/3k^2\Phi - k^2 c_s^2\Theta,$$

so this equation of a forced harmonic oscillator, and one can show the equation and find out how the acoustic peaks of different multipoles will behalf, which is summarized in the figure. Note that:

(1) the amplitude is sensitively dependent on baryon density, not surprisingly. This is mostly due to the slowing down of sound wave with the presence of baryon.

(2) the location of the peak depends mostly on curvature, with little dependence on mass density and very little on cosmological constant.

(3) The first peak, which actually corresponds to half wave to recombination, is the highest, second is lower. Then, high order peaks are suppressed due to various small scale effect that we are going to discuss now.

7 Small Scales

Small scale effects include:

- statistical and silk damping. Both statistical damping due to the finite thickness of last scattering layer, which will result in statistical reduction for small scale fluctuations. SIlk damping is the photon diffusion effect before recombination and decoupling, which will also suppress small scale power. Both has scale of 10Mpc, or about a few arcmin, resulting in damping at $l > 500$, with severe damping at $l > 2000$.
- The Sunyaev-Zeldovich effect. We will discuss S-Z later. Angular scale of high-redshift clusters is about 1 arcmin. Hot plasma in the cluster will scatter photons. Since electrons are hotter, in this process photons will gain energy. In the microwave sky, which is on the R-J tail of black body, this will result in a reduction in temperature. There is also a kinematic effect caused by the peculiar velocity of clusters. They need to be modeled and taken into account for high-l.
- Confusion due to discrete sources. In cm wavelength, discrete sources are mostly quasars, and also radio galaxies, BL Lacs. For high frequency observations, such as those with Planck, small scale power from clustering of sub-mm galaxies can affect things as well.

8 Reionization

At some epoch well after the epoch of recombination, the IGM gas must have been heated and reionized. We will discuss this in great detail in a few weeks. This is the process of reionization, when the universe changed from HI to HII. Figure.

- When: z=6-20; latest Planck, $z \sim 8 - 9$.
- By what: most likely high-energy photons from early galaxies.
- how: not clear at all

From CMB point of view, because reionization released hot photons, there will be a finite τ for Thompson scattering. The next effect of this to attenuate temperature fluctuation and therefore the power spectrum on scales less than reionization horizon size.

Figure. Note that $\tau = 0.1$ and 0.2 correspond to $z = 12$ and 20. This is probably a bit too high.