Chapter 8

The universe: size, origins, contents

8.1 Introduction

The general and special theories of relativity discussed in the previous chapters are the tools currently used in the investigation and description of the universe. Most of the objects in the universe are somewhat mundane: stars, planets, rocks and gas clouds. Yet in many respects the universe is far from being a placid and peaceful place. There are stars which explode with the energy of a billion suns, black holes with millions of times the mass of our sun which devour whole planetary systems, generating in one day as much energy as our galaxy puts out in two years. There are enormous dust cluds where shock waves trigger the birth of new stars. There are intense bursts of gamma rays whose origin is still uncertain.

These phenomena are not infrequent, but appear to be so due to the immense distances which separate stars and galaxies; for one of the most impressive properties of the universe is its size. The universe is so large that just measuring it is very difficult, and finding out the distance to various objects we observe can be a very complicated proposition.

In order to extract information about the universe a toolbox of methods has been devised through the years. I will first discuss the most important of these methods, and with these I will describe how measure the universe and discuss its evolution. We need to determine sizes and distances because, as we will see, they provide basic information about the history of the universe.

Most of the data we get from the universe comes in the form of light (by which I mean all sorts of electromagnetic radiation: from radio waves to gamma rays). It is quite remarkable that using only the light we can determine many properties of the objects we observe, such as, for example, their chemical composition and their velocity (with respect to us). In the first two sections below we consider the manner in which we can extract information from the light we receive.

But detecting light is not the only way to obtain information from the universe, we also detect high-energy protons and neutrons (forming the majority of cosmic rays). The information carried by these particles concerns either our local neighborhood, or else is less directly connected with the sources: isolated neutrons are not stable (they live about 10 minutes), so those arriving on Earth come from a relatively close neighborhood (this despite time dilation - Sect. **??**). Protons, on the other hand are very stable (the limit on their lifetime is more than 10^{32} years!), but they are charged; this means that they are affected by the magnetic fields of the planets and the galaxy, and so we cannot tell where they came from. Nonetheless the more energetic of these particles provide some information about the most violent processes in the universe.

In the future we will use yet other sources of information. Both gravitational wave detectors and neutrino telescopes will be operational within the next few years. Neutrinos are subatomic particles which are copiously produced in many nuclear reactions, hence most stars (including our Sun) are sources of neutrinos. These particles interact very very weakly, and because of this they are very hard to detect. On the other hand, the very fact that they interact so weakly means that they can travel through very hostile regions undisturbed. Neutrinos generated in the vicinity of a black hole horizon can leave their native land unaffected and carry back to Earth information about the environment in which they were born.

8.2 Light revisited

In this section I will describe two properties of the light we receive and the manner in which it can be used to extract information about its sources.

8.2.1 The inverse-square law

A source of light will look dimmer the farther it is. Similarly the farther away a star is the fainter it will look; using geometry we can determine just how a star dims with distance

Imagine constructing two spheres around a given star, one ten times farther from the star than the other (if the radius of the inner sphere is R, the radius of the outer sphere is $10R$). Now let us subdivide each sphere into little squares, 1 square foot in area, and assume than on the inner sphere I could fit one million such squares. Since the area of a sphere increases as the square of the radius, the second sphere will accommodate 100 times the number of squares on the first sphere, that is, 100 million squares (all 1 square foot in area). Now, since all the light from the star goes through both spheres, the amount of light going through one little square in the inner sphere must be spread out among 100 similar squares on the outer sphere. This implies that the brightness of the star drops by a factor of 100, when we go from the distance R to the distance $10R$ (see Fig. 8.1).

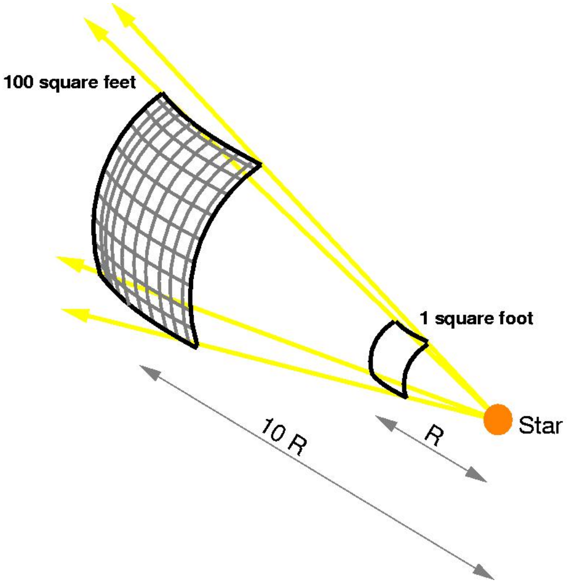

Figure 8.1: Illustration of the inverse-square law: all the light trough the 1 square-foot first area goes through the second one, which is 100 times larger, hence the light intensity *per square foot* is 100 times smaller in the second area. The intensity drops as $1/R^2$.

If we go to a distance of $20R$ the brightness would drop by a factor of 400, which is the square of 20, for $30R$ there would be a decrease by a factor of $900 = (30)^2$, etc. Thus we conclude that

The brightness drops as $1/\left(distance\right)^2$.

Light intensity drops as $1/\left(\text{distance}\right)^2$.

This fact will be used repeatedly below.

8.2.2 The Doppler effect

We have seen that light always travels at the same speed of about 300, 000km/s; in particular light emitted by a sources in relative motion to an observer travels at this speed. Yet there is one effect on light which shows that its source is moving with respect to the observer: its color changes.

Imagine standing by the train tracks and listening to the train's horn. As the train approaches the pitch of the blast is higher and it becomes lower as the train recedes from you. This implies that the frequency of the sound waves changes depending on the velocity of the source with respect to you, as the train approaches the pitch is higher indicating a higher frequency and smaller wavelength, as the train recedes from you the pitch is lower corresponding to a smaller frequency and a correspondingly larger wavelength.

This fact, called the *Doppler effect*, is common to *all* waves, including light waves. Imagine a light bulb giving off pure yellow light; when it moves towards you the light that reaches you eye will be bluer, when the bulb moves away form you the light reaching your eye will be redder. If you have a source of light of a known (and pure) color, you can determine its velocity with respect to you by measuring the color you observe. Qualitatively, if one observes a redder color (longer wavelength than the one you know is being emitted) then the source is moving away from you, if bluer (shorter wavelength that the one you know is being emitted) the source is moving toward you (see Fig. 8.2).

If one observes a redder color (longer wavelength than the one you know is being emitted) then the source is moving away from you, if bluer (shorter wavelength that the one you know is being emitted) the source is moving toward you

The important point here is that knowing the frequency at the source and measuring the observed frequency one can deduce the velocity of the source [1] If the source is moving sufficiently fast towards you the yellow light will be received as, for example, X-rays; in this case, however, the source must move at 99.99995% of the speed of light. For most sources the shift in frequency is small.

8.2.3 Emission and absorption lines

When heated every element gives off light. When this light is decomposed using a prism it is found to be made up of a series of "lines", that is, the output from the prism is not a smooth spectrum of colors, but only a few of them show up. This set of colors is unique to each element and provides a unique fingerprint: if you know the color lines which make up a beam of light (and you find this out using a prism), you can determine which elements were heated up in order to produce this light.

---

[1] More precisely this is the velocity along the line of sight,

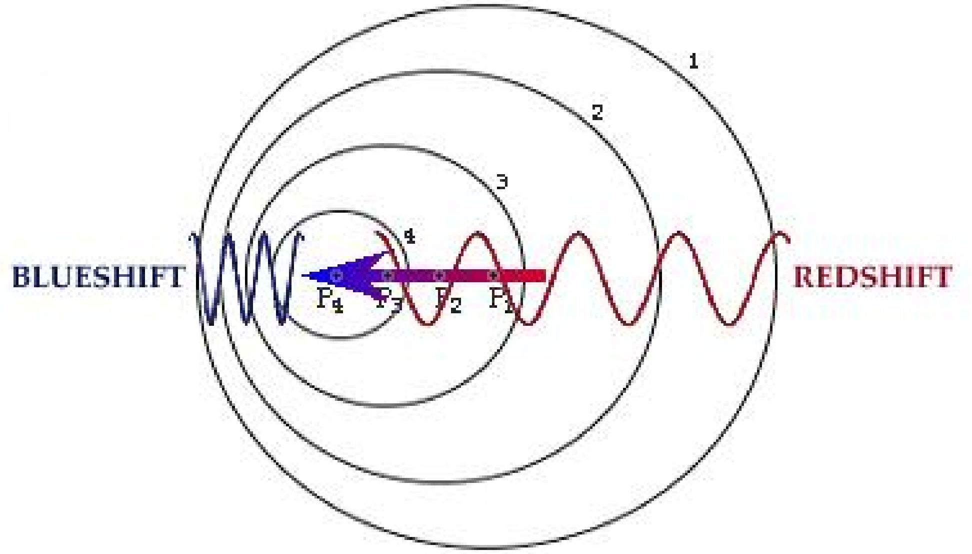

Figure 8.2: Diagram illustrating the Doppler effect. The source is moving to the left hence a receiver on the right will see a red-shifted light while a receiver on the left will see a blue-shifted one. .

Similarly, when you shine white light through a cold gas of a given element, the gas blocks some colors; when the "filtered" light is decomposed using a prism the spectrum is not full but shows a series of black lines (corresponding to the colors blocked by the gas); see Fig. 8.3. For a given element the colors blocked when cold are exactly the same as the ones emitted when hot.

The picture in Fig. 8.3 corresponds to a single element. For a realistic situation the decomposed light can be very complex indeed, containing emission and absorption lines of very many elements. An example is given in Fig. 8.4.

Each element presents a unique set of lines

After the discovery of emission and absorption lines scientist came to rely heavily on the fact that each element presents a unique set of lines: it is its inimitable signature. In fact, when observing the lines from the solar light, it was found that some, which are very noticeable, did not correspond to any known element. Using this observation it was then predicted that a new element existed whose absorption lines corresponded to the ones observed in sunlight. This element was later isolated on Earth, it is called Helium (from *helios*: sun).

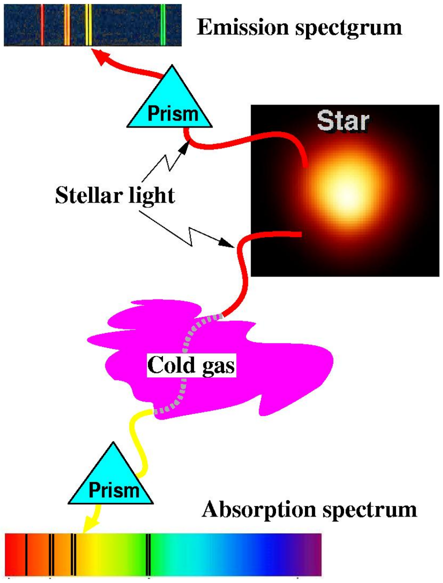

Figure 8.3: Diagram illustrating emission and absorption lines: when light given off by hot gas is decomposed using a prism it is shown to be made up of colored lines (emission lines). When white light shines trough a cold gas the resulting light , when decomposed is shown to have dark lines (absorption lines). The emission and absorption lines for the same element match.

In following this line of argument one has to be very careful that the lines are not produced by *any* other element. This is complicated by the fact that some lines are observable only under extreme circumstances and one has to take them into consideration as well. For example, after the success of the discovery of Helium, *another* set of lines (not so prominent) was isolated and associated with yet another element, “coronium”. It was later shown that the coronium lines were in fact iron lines, which are clearly observable only in the extreme conditions present in the sun (one can also see them in the laboratory, it’s just hard to do so).

8.2.4 A happy marriage

When observing stellar light from various distant stars (decomposed using a prism) it was found that, just as for the sun, they presented lines. But, curiously, these lines corresponded to no known element! This may imply that each star carries a new set of elements, but the simplest hypothesis (which should be investigated first, see Sect. **??**) is that the mismatch between the

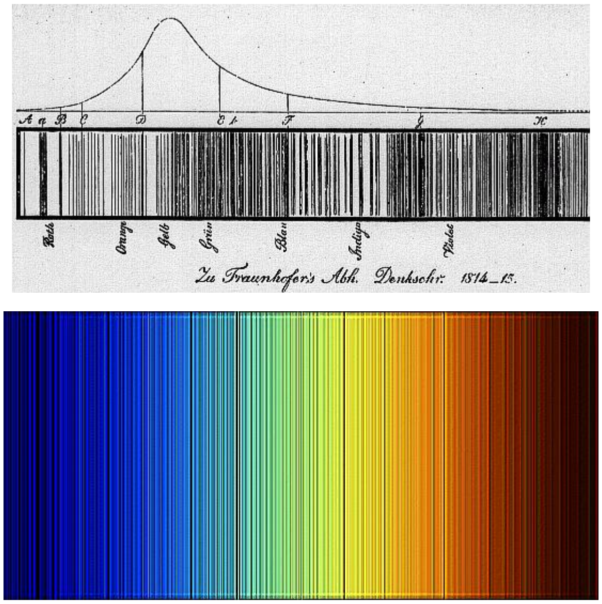

Figure 8.4: Solar light decomposed by a prism exhibiting the emission and absorption lines. At the top is one of the first of such measurements (1817); the curve above the lines denotes the intensity of the various colors, as expected it is largest in the yellow. The second figure is a modern photograph of the solar absorption lines.

laboratory and stellar lines is due to the Doppler effect which will shift the lines towards the red or blue according to the motion of the star (which is the source in this case) with respect to Earth. One can then use the shift in the observed stellar lines to determine the velocity at which the star is moving (with respect to us) and also the elements in it. In one fell swoop we determine the constitution and the speed of the stars using only the light we get from it.

Using spectral lines we can determine both the speed of the star and the elements in it

8.3 Cosmic distance ladder

Another important piece of information regarding objects in the universe is their distance to us. This is not an easy thing to measure since these objects are usually very far apart. I will measure distances in light years: one light year is the distance covered by light during one year, which is about 9.5 trillion kilometers, or about 6 trillion miles.

In order to understand why several steps are needed in measuring distances it is useful to consider a simple example. A student is in her room sitting at her desk and would like to find the distance to the window; she gets a ruler and laboriously measures this distance to be 3 feet. This I will call "the first rung in the student's distance ladder"

Her next task is to find the distance to a building which she can see through the window. This building is too far away for her to use her 12 inch ruler. What she does is to use sound: she notices that when she claps her hands outside her window there is an echo produced by the sound bouncing off the building in front of her. She has a good watch and so she can determine the time it takes for the sound to get from her window to the building and back. Now, if she can determine the speed of sound, she could use the formula $distance = speed \times time$ to get the distance. In order to measure the speed of sound she closes her window and times the echo from her desk to the window. Since we already knows the distance to the window (which she measured using her ruler) and she now knows the time it takes sound to go from her desk to the window and back she can determine the speed of sound. So, *using the first measurement* she determines the speed of sound and this allows her to measure things that are much farther away. In this way she has "constructed" the second rung in her distance ladder.

The same idea is used when measuring far away things in space: one finds a reliable method to determine the distances to near-by stars (the equivalent of using the ruler). Then one devises another method which requires a sort of calibration (the equivalent of determining the speed of sound); once this calibration is achieved the second method can be used to find distances to objects that are outside the range of the first method. Similarly a third, fourth, etc. methods are constructed, each based on the previous ones.

Step 1: distances up to 100 $\ell.y.$

For near-by stars their distance is measured by parallax: the star is observed in, say, December and then in June, and the direction of the star with respect to the sun is measured in both cases. Knowing these angles and the diameter of the orbit of the Earth around the sun, one can determine the distance to the star (see Fig. 8.5).

As we look at farther and farther stars the angles measured come closer and closer to 90^o. For stars more than 100 $\ell.y.$ from Earth one cannot distinguish the angles from right angles and the method fails.

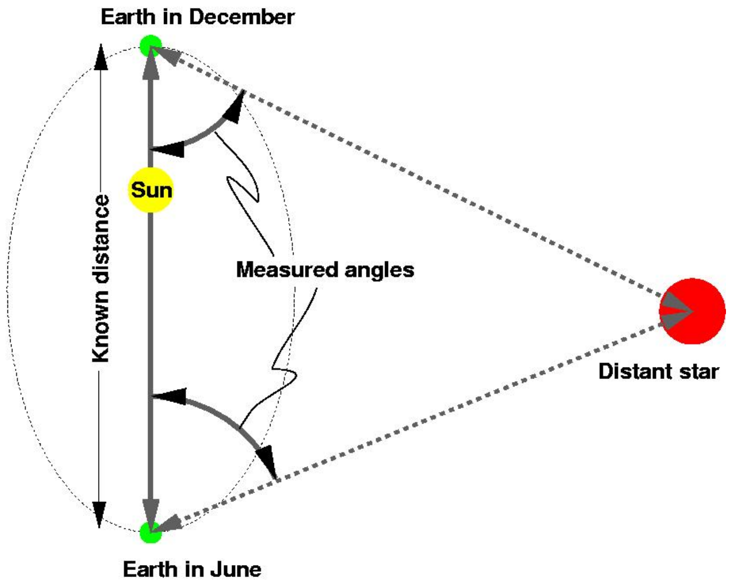

Figure 8.5: Knowing the size of Earth's orbit and measuring the angles of the light from the star at two points in the orbit, the distance to the star can be derived. The farther the star is, the smaller the angles.

Step 2: distances up to 300,000 $\ell.y.$

In the decade 1905-1915 Hertzprung and Russel observed a group of near-by stars whose distances they knew (using parallax). For each star they recorded its color and calculated its brightness as it would be measured at a distance of 1 $\ell.y.$ (using the $1/(\text{distance})^2$ law, see Sect. 8.2.1). Then they plotted this brightness versus the color; what they found is that most stars (90% of them) lie on a narrow band in this type of plot which they called the *main sequence* (see Fig. 8.6).

Suppose we now obtain the HR plot for stars which are far away, say on the other side of the galaxy, about 10^5 light years (10^{18} km). If we choose these stars such that they are not too far apart (there are good astronomical indicators for this) the distance from Earth to any such star will be more or less the same. It is found that, as for the near-by stars, 90% of these far stars will again fall on a main-sequence strip in the color vs. brightness plot.

On the other hand all these stars are dimmer than the near-by stars originally used by H&R; the decrease in brightness is due to the fact that brightness drops as the square of the distance (Sect 8.2.1). Comparing the two main sequences (for near and far stars) as in Fig. 8.7, we can extract

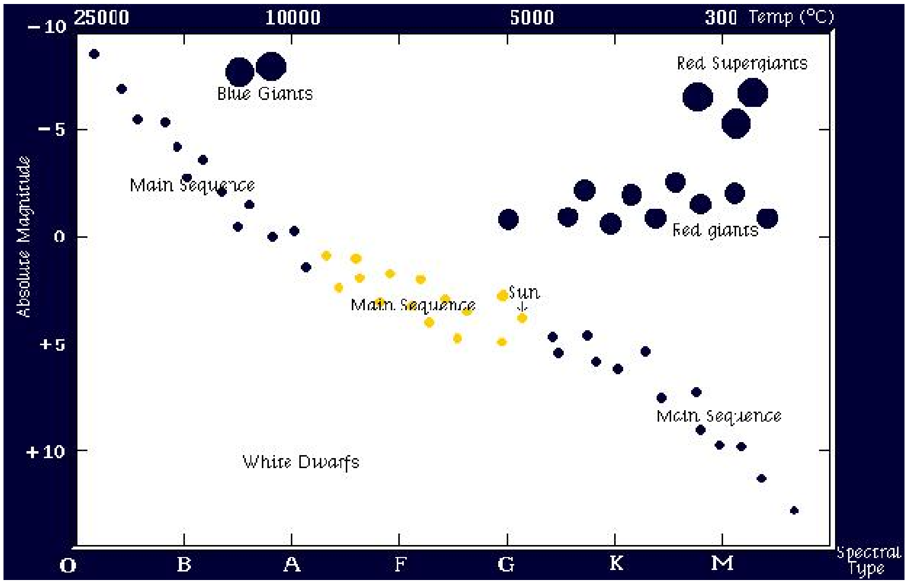

Figure 8.6: The Hertzprung-Russel diagram. The horizontal axis corresponds to the color of the star: blue to the left, red to the right. The vertical axis corresponds to the star's brightness (brighter stars are plotted higher). Though the diagram does not represent it, the groups labeled red supergiants, red giants, blue giants and white dwarfs, are but a small fraction of the whole stellar population, most stars are in the main sequence.

the distance to these far-away stars. This method can be used to determine distances up to 300,000 $\ell.y.$; for larger distances the main sequence stars are too dim to obtain a reliable estimate of their brightness.

Step 3: distances up to 13,000,000 $\ell.y.$

In 1912 Henrietta Swan Leavitt noted that 25 stars, called Cepheid stars [2] (their location in the HR diagram is given in Fig. 8.10), in the Magellanic cloud [3] (see Fig, 8.8) are variable, that is, they brighten and dim periodically. Many stars are variable, but the Cepheids are special because their period (the time they for them to brighten, dim and brighten again, see Fig. 8.9) is

---

[2] The name derives from the constellation in which they were first observed.

[3] This is a small galaxy (of only 10^8 stars) bound to the Milky Way.

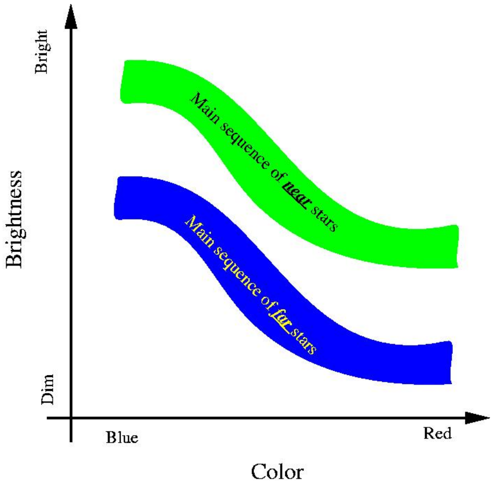

Figure 8.7: The Hertzprung-Russel diagram for the main sequence of near and far stars. The comparison is used to determine the distance to the far stars.

i) regular (that is, does not change with time), and

ii) a uniform function of their brightness (at a 1 light-year distance). That is, there is relation between the period and brightness such that once the period is known, the brightness can be inferred.

Leavitt was able to measure the period by just looking at the stars and timing the ups and downs in brightness,

But in order to obtain the brightness at the distance of one light year she needed to fist measure the maximum brightness on Earth and then, using the HR method, determine the distance from Earth to these stars (as it turns out, the Magellanic cloud is about 10^5 light years away from us).

What she obtained is that the brighter the Cepheid the longer its period, and that the relation between brightens and period was very simple: a straight line (Fig. 8.11). This means that the period and brightness are proportional to each other

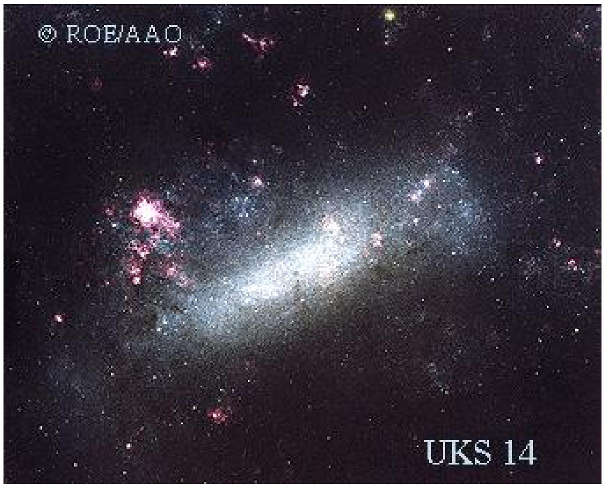

Figure 8.8: The Magellanic Cloud

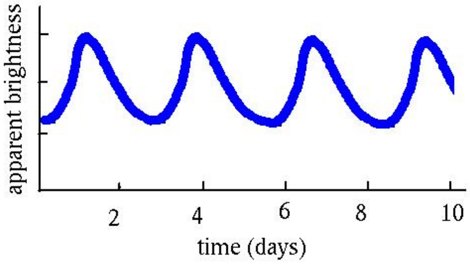

Figure 8.9: Illustration of the brightening and dimming of a variable star.

Measuring properties of the Cepheid variables. The color is no problem, you just observe the starlight through different color filters and observe the intensity; 'the' color of the star corresponds to the filter which lets pass the highest intensity light. The intensity at the distance of one light year is obtained by measuring the intensity on Earth and calculating the distance to the star, then one uses the fact that the intensity drops as the square of the distance. For example, suppose we observe a star which has intensity of 1 (in some units), and which we know is at a 10 $\ell.y.$ from Earth, then at a distance of 1 $\ell.y.$ (which is 10 times smaller) the intensity will be 100 times larger (the square of 10 is 100) and so the intensity at the distance of 1 $\ell.y.$ will be 100 (in the same units as before).

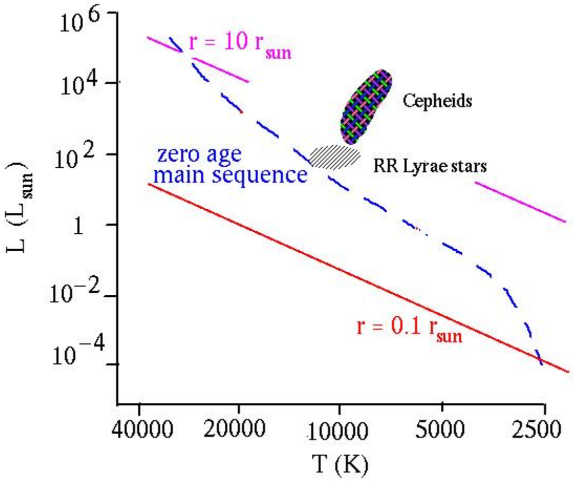

Figure 8.10: The location of the Cepheid variable stars in the Hertzprung-Russel plot.

These stars are quite distinct, reasonably abundant and very bright. One can identify them not only in our galaxy, but in many other galaxies as well.

If one requires the distance to a given galaxy one first locates the Cepheid variables in this galaxy. From these observations one determines the period of each of these stars. Leavitt's data states that a given period has a unique brightness associated to it. So form the period and Leavitt's plot we get the brightness at the distance of one light year. We can also measure the brightness on Earth. The brightness at the distance of one light year will be larger than the observed brightness due to the fact that this quantity drops like the square of the distance (Sect. 8.2.1). From these numbers one can extract the distance to the stars. This method works up to 13 million $\ell.y.$ when Earth-bound telescopes are used; for larger distances these stars become too dim to be observed.

Much more recently the Hubble telescope has used this same type of indicators to much farther distances (the Hubble is outside the Earth's atmosphere and can detect much fainter stars). Looking at a galaxy in the Virgo cluster (the galaxy is "called" M100), Wendy L. Freedman found (1994-5) that the Cepheid variables in this galaxy could be used to determine its distance; the result is 56 million $\ell.y.$.

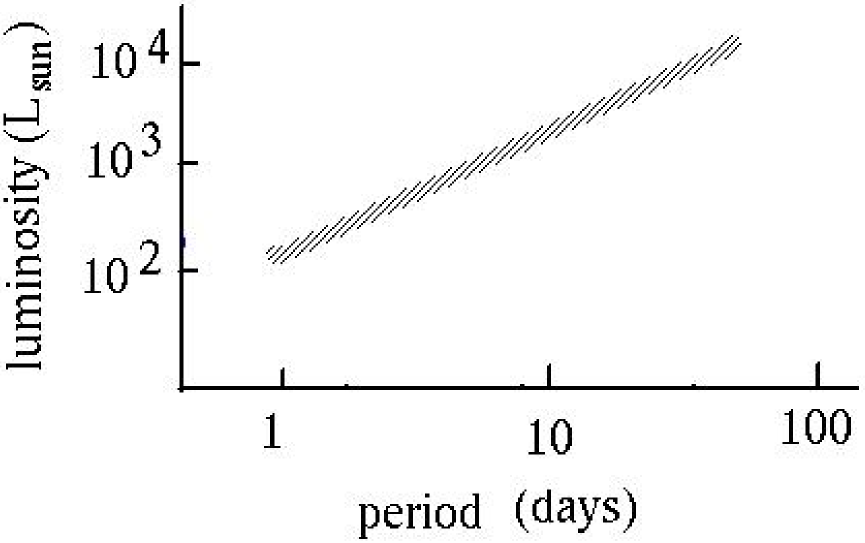

Figure 8.11: Relationship between the brightness and period of the Cepheid variable stars.

Young Cepheids. Recent observations of Cepheid variables in the galaxy M100 from the Hubble telescope have generated some puzzling questions. Using these observations (and the General Theory of Relativity) it is possible to determine the age of the universe: we measure both the distance and the velocity of these objects (with respect to us) and we can calculate the rate of expansion of the universe, from this we get the time it took to get to its present size. Curiously enough the age is in conflict with some other age determinations: some stars are older than the number obtained!

How can this be resolved? There are several possibilities. one of the most likely ones is that, since the Cepheid is observed by the Hubble telescope are very far away, the light we get was sent out when the stars were quite young. But it has not been shown that Leavitt's data is also valid for such teenage stars. It is quite possible that these stars have a different behavior and only settle into regular predictability only as they become middle-aged.

Step 4: distances up to 1,000,000,000 $\ell.y.$

For larger distances run of the mill stars are of no use: they are too dim. There are, however, some stars which at the end of their life blow themselves apart and, in doing so, become anomalously bright (out-shining a galaxy in many cases) for a brief period of time (less than a month); such an object is called a supernova (for more details see Sect. **??**). The unique characteristics and enormous brightness of a certain type of supernova can be used to determine distances beyond the reach of the previous methods.

There have been many measurements of the manner in which a supernova, whose distance to Earth is known (using one of the previous methods), increases its brightness and then dims into oblivion. There is one type (called type Ia) for which this brightening and dimming is very regular: when the maximum brightness at a distance of 1 $\ell.y.$ is calculated (using the known distance and the $1/\text{distance}^2$ rule), it is found to be the same for all cases [4].

If the distance to a far away galaxy is required, one must first locate a type Ia supernova in it (which do occur regularly) and then measure its observed brightness. Comparing this result with the known maximum brightness (at a $1\ell.y.$ distance) achieved by all such supernovae one can determine the distance to the galaxy in question (again using the $1/\text{distance}^2$ rule). Since supernovae are extremely bright this method is useful to very large distances, up to 10^9 $\ell.y.$.

Step 5: distances beyond 1,000,000,000 $\ell.y.$

For very far objects none of the above methods work. The reason is interesting: since we are looking at very distant objects their light has taken a very long time to reach us, so the light we get must have left the object a long time ago. Because of this the farther we look the earlier the images we get: looking far away is equivalent to looking back in time. When we look at the farthest obects we can see, what we get are images of their early stages of their development.

In addition, since the brightness drops as the square of the distance, these far objects must also be very bright. From this it follows that the most distant objects we see are necessarily very bright and very young.

In order to determine the distances with any degree of accuracy we need to know the brightness at a distance of 1 $\ell.y.$, but here we hit a stone wall: the only objects we see are much older than the ones we are interested in,

---

[4] In doing so astronomers must select type Ia supernovae that exhibit no abnormalities, else the measurements might be corrupted.

and we do not have a reliable theory of the way in which these things evolve, we have no way of calibrating our observations using any near-by objects.

It is here, in the observation of the universe at large, that the General Theory of Relativity must be used to measure distances. How this is done is described in the next section.

8.4 The relativistic universe

In everyday life there are many forces that strongly affect the world around us: friction, electric, magnetic, etc. But in the universe at large there is only one predominant force: gravity. It is gravity that determines the structure of the universe at large.

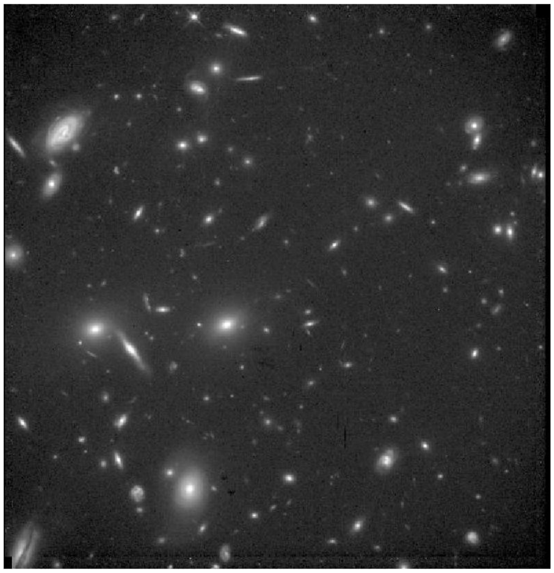

Figure 8.12: NASA Hubble Space Telescope image of the central portion of a remote cluster of galaxies (CL 0939+4713).

The (visible) universe is filled with galaxies (see Fig.8.12) each containing a billion suns (more or less) tightly bound by their mutual gravitational atraction. Because of this we can think of a galaxy as a solid object of a given mass (in the same way that when you look at the gravitational pull of the Earth on the Moon you don't have to worry about the fact that they are made of atoms; the stars are the "atoms" which make up galaxies).

Magnitudes. The typical galaxy like the Milky Way has most of its stars in a central bulge of 10^4 $\ell.y.$ diameter or less, where about 10^{11} sun-sized stars are concentrated. The pull of these stars on out Sun is 10^4 times stronger than that of the nearest sizable galaxy (Andromeda) which is at about 2×10^6 $\ell.y.$ away and also has about 10^{11} sun-sized stars.

In this simplified picture the visible matter in the universe (that which shines) is concentrated in a dusting of galaxies. In addition the universe can contain matter which does not shine, such as planet-sized objects, cold dust and, perhaps, other more exotic objects (see Sect. 8.5.1). The universe also contains electromagnetic radiation: for example, stars continuously give off light and heat (infrared radiation) which then disperses throughout the universe (this is why we can *see* them!). Finally the universe contains a significant amount of microwaves (see Sect. 8.4.2) and neutrinos, (see Sect. 8.5.1), both relics from a very early time.

The first person to look at the cosmos through the eyes of the General Theory of Relativity was Einstein himself. He took the above picture of a universe filled with matter and radiation he added two assumptions

- *Homogeneity:* on average the universe looks the same from every vantage point.
- *Isotropy:* on average the universe looks the same in every direction

These assumptions, though reasonable, still require justification; I will come back to them. With these preliminaries one can solve the equations of the General Theory of Relativity and find a description of the universe and the manner in which it evolves.

To Einstein's initial surprise there were no steady solutions: the universe according to the General Theory of Relativity *must* expand or contract. He compared this result with the best observational data of the time and found, to his dismay, that the observations strongly favored a steady universe. He then made what he called "the greatest scientific blunder of my life": he modified the equations of the General Theory of Relativity by adding a term that countered the expansion or contraction present in his initial solutions [5]. With this *ad hoc* modification he did find a steady universe and was (temporarily) satisfied.

---

[5] The modification amounts to the inclusion of a uniform cosmic pressure which balances the tendency to the universe to expand.

Not long afterwards Hubble published his now famous observations that demonstrated that our universe is, in fact, expanding; and the manner in which it expands agrees with the predictions of the solutions first obtained by Einstein. It was then that Einstein, to his satisfaction, dropped his modification of the equations. But this was not the end of this saga: the added term, like the genie from the bottle, refused to disappear, showing up in many models (recent observations suggest that it must be included in order to account for the observations). I will come back to this in Sect. 8.5.2.

What Hubble did was to measure the red-shift of a group of galaxies whose distances he knew (there were no blue-shifted galaxies, which means that these galaxies were receding from the Milky Way). Using the measured red-shift and the formulas for the Doppler effect, he found the speed at which they receded. Then he made a plot (called now a "Hubble plot") of velocity vs. distance and found that, *as predicted by the General Theory of Relativity* all points fall in a straight line (see Fig. 8.13); the slope of this line is called *Hubble's constant*. General Relativity then predicts that the distance d to an object is related to its velocity v (both measured with respect to the Earth) by

$$v = H_o\, d$$

which is called *Hubble's law* and H_o is Hubble's constant, its value is approximately

$$H_o = \frac{1}{1.5 \times 10^{10}\text{years}}.$$

General Relativity then predicts that the distance d to an object is related to its velocity v both measured with respect to the Earth by $v = H_o\, d$ which is called *Hubble's law* and H_o is Hubble's constant

It is the above relation between distance and velocity that is used to measure distances beyond 10^9 $\ell.y.$: the final step in the cosmic distance ladder. Needless to say astronomers have verified Hubble's law for distances below 10^9 $\ell.y.$ using supernovae (Sect. 8.3). In order to find the distance to the farthest objects in the universe one first obtains their redshift and, using Doppler's formulas, derives the velocity v of the object. The distance is then v/H_o.

8.4.1 The expanding universe

All of Hubble's (and subsequent) measurements indicate hat all galaxies are receding from the Milky way and its neighbors. One might think that we are being ostracized by the universe as a whole, that the Milky Way has become a cosmic pariah; but a little thought shows that this is not the case. According to General Relativity the universe is expanding, but this does

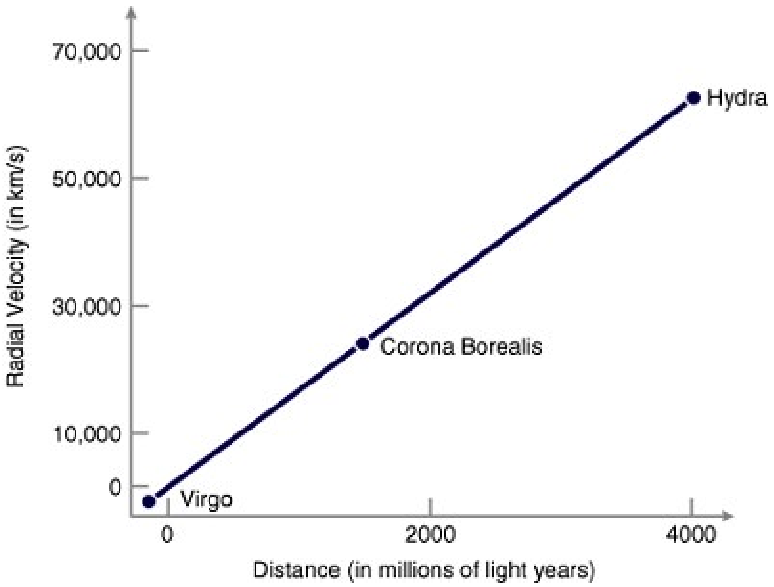

Figure 8.13: Illustration of Hubble's law.

not mean that the galaxies and such are flying out into space, it means that *space itself* is growing, and in so doing, it increases the separation between the galaxies. The classical example is to imagine a balloon with dots drawn on it; the balloon's latex represents space, the dots represent the galaxies. As the balloon is inflated (space grows) the distance between the dots (the galaxies) increases. An observer in any one dot would see the other dots receding from him/her (just as we see distant galaxies receding from us).

This universal expansion represents only the average motion of the galaxies, the motion of a given galaxy can present deviations from this average. For example, galaxies which are close together are bound by their mutual gravitational pull and this distorts the Hubble flow.

The General Theory of Relativity predicts that the universe is not static, and observations confirm this indicating that it is expanding. Thus the universe must have been smaller in the past, and, following this idea to its limit, must have been a point in its inception. Thus the universe began at a point, in the distant past and has been expanding ever since. The event marking this beginning is known (with a characteristic scientific flair for words) as the *Big Bang*.

Just after the Big Bang the universe contained an extremely hot and dense soup of matter and energy (which are equivalent in the sense of the

Special Theory of Relativity) under which conditions any kind of object would melt almost instantaeously into its components. Yet the universe expanded and cooled accordingly, and this cooling allowed for the formation of more and more complicated structures, ranging from atoms ($300,000$ years after the Big Bang) to Galaxies (10^9 years after the Big Bang) (see Fig. 8.16).

It must be remembered that the Big Bang represent the creation of the universe, *including* space and time. The Big Bang is *not* to be pictured as a big explosion somewhere out in space with galaxies being spewed out from the explosion region. Instead the picture provided by General Relativity is of the whole universe, including space, appearing at the Big Bang and expanding after that (like the balloon model described above). In this picture the Big Bang occurred *everywhere*.

The Big Bang occurred *everywhere*.

And now what?

The universe expanding, but what will become of it? There are three possible solutions to the equations of the General Theory of Relativity which represent homogeneous and isotropic universes: either it will continue its expansion forever, or it will eventually stop and re-contract or it will expand slowing down to a stop at infinite time. The contents of the universe (matter and radiation) determine which of these is realized in *our* universe. In all three cases the shape of space remains the same as the universe expands (or in the second case, as it expands and contracts).

The universe will continue its expansion forever, or it will eventually stop and re-contract or it will expand slowing down to a stop at infinite time.

That the shape of space is determined by the amounts of matter and energy in the universe is not surprising as it is matter and energy which determine the curvature of space (see Sect. **??**).

- Space in an eternally expanding or *open* universe is shaped like a 3-dimensional horse saddle. In this case the angles in a triangle add up to *less* than 180^o.
- Space in a *closed* universe which will eventually re-contract is shaped like a 3-dimensional sphere. In this case the angles in a triangle add up to *more* than 180^o.
- Space in a *flat* universe which expands slowing down to a stop at infinite time is shaped like a 3-dimensional plane. In this case the angles in a triangle add up to 180^o.

Space in an eternally expanding or *open* universe is shaped like a 3-dimensional horse saddle

Space in a *closed* universe which will eventually re-contract is shaped like a 3-dimensional sphere

Space in a *flat* universe which expands slowing down to a stop at infinite time is shaped like a 3-dimensional plane

These possibilities are illustrated in Fig. 8.14.

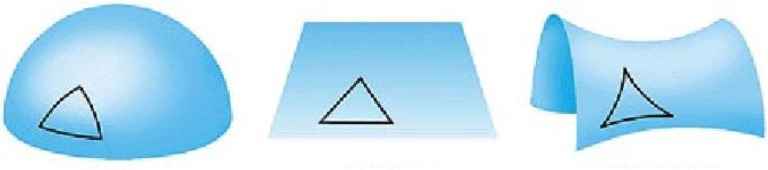

Figure 8.14: The three possible shapes of a homogeneous and isotropic universe: a closed universe (left), a flat universe (center) and open universe (right). See the text for an explanation.

These three possibilities give the *average* shape of space. Individual masses produce local bumps and troughs. This is similar to the way we talk about the Earth: we say it is a sphere, though we know it is full of bumps (for example, Himalayas) and troughs (the Dead Sea, for example).

Of these possibilities the one corresponding to our universe is determined by the amount of matter in the cosmos. If there is very little the initial thrust from the Big Bang will never be stopped, if however there is a large amount of matter, the mutual gravitational pull will be sufficient to break the expansion and eventually cause a re-contraction. Hence there is a critical amount of matter such that if our universe has more it will re-contract, if less it will expand forever (if it has precisely the critical amount it will expand forever slowing down to a stop at infinite time). These possibilities are illustrated in Fig. 8.15.

The obvious question is then: how much stuff is in the universe? And to that we can say: we don't know. If we count all the matter that shines (stars and such) we get a number very low compared to the critical value. But, is most of the matter shining? Could it not be that there is a lot of dust out there? The latest results suggest that the universe will expand forever, but at present its ultimate fate is unknown.

8.4.2 The Microwave Background Radiation

General Relativity not only provides a nice history of the universe, but it also points out viable measurements which can support its validity. The most important is the so-called *Microwave Background Radiation*.

When the universe began the density and temperature of the initial fireball was so high that all matter dissociated into its primary components. Note also that in this initial setting the force of gravity was enormous. As the expansion progressed the universe cooled and the initial fundamental

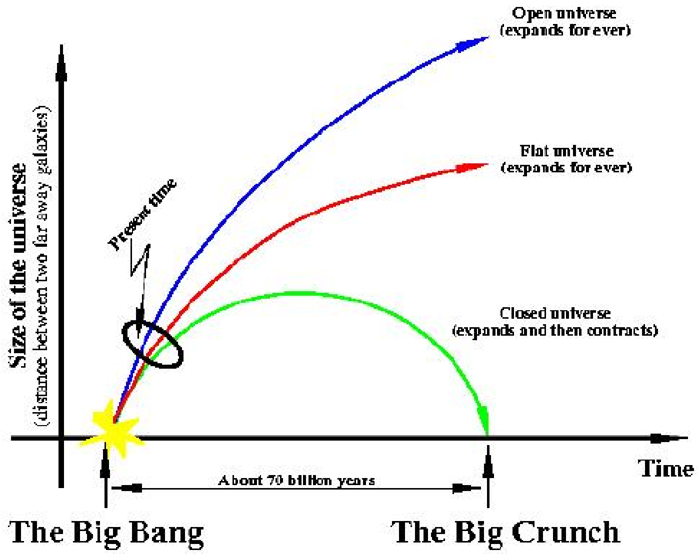

Figure 8.15: The universe might expand forever or will re-contract

constituents formed increasingly more complicated objects. This is so because when the temperature is very high everything is jiggling very fast and anything that can be dissociated will; as the temperature drops so does the jiggling and, eventually, composite structures can form and survive. Thus, if we had been able to film the contents of the universe as it cooled, and then run the film backwards we would first see atoms which are then broken apart into nuclei and electrons by the intense heat, then we would see the nuclei themselves decomposing into protons and neutrons, then the protons and neutrons decomposing into quarks [6]. The microwave background radiation is a messenger from this primordial soup.

To understand how why is this microwave radiation present and how it was generated I need to talk a bit about the way charged bodies interact with light. Remember now that light is described by the same equations that describe the physics of electric charges (Maxwell's equations), this suggests (and it is true) that light will interact with charged objects. In fact this is how your skin gets hot when exposed to the sun: your skin is composed of molecules which are made of atoms. Atoms in their turn are composed of a small heavy nucleus (with positive charge) surrounded by a cloud of

---

[6] There are many hypotheses about the way the universe looked at times before that of quark formation, but none has been accepted yet this is an area of active research.

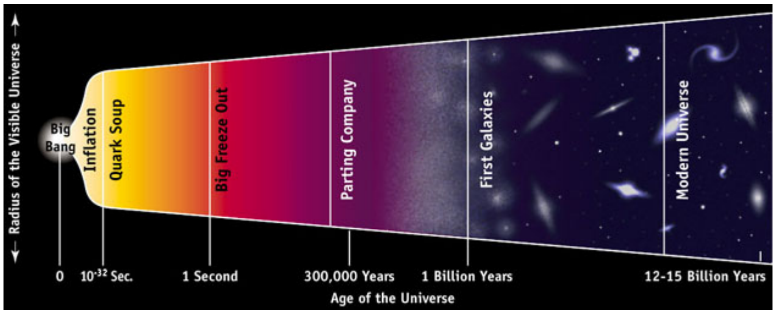

Figure 8.16: Abbreviated history of the universe according to the Big Bang model.

negatively charged light particles, the electrons. When light shines on your skin it is absorbed by the electrons which get agitated, and it is this agitation which you perceive as heat. This is not as efficient as it might be because the electrons are not free, they are inside atoms, so that *on average* the atoms are neutral. Much more light would be absorbed by a set of free electrons. This also works in the reverse: if you jiggle electrons sufficiently rapidly they will give off light, this is how a light-bulb works.

Suppose now that you have a box with perfectly reflecting walls and which are kept very hot. Into that box we introduce a bunch of electrons and nuclei and also light. Assume that the system is so hot that the electrons are not bound to the nuclei: as soon as they come close they are wrenched apart by the intense heat of the environment. So, on average, what you see is a bunch of charged particles and light running amok. In this case light is constantly being absorbed and emitted by the electrons and nuclei.

Now imagine that you cool the box by making it larger. Eventually things will get cold enough for the electrons to stay attached to the nuclei, the heat is not sufficiently high for them to be wrenched apart. At this point the rate at which light is absorbed and emitted drops rather suddenly for now the particles in the box are neutral (on average). From this point on light will just stream forth unimpeded (until it is reflected by a wall).

This is precisely what happens in the universe. After the big bang there came a point where electrons and nuclei were formed. They were immersed in intense electromagnetic radiation (light, X-rays, gamma rays, etc.). As time progressed and the expansion of the universe continued, the system became cooler (much as for the box when we increased its size). Eventually a point was reached where the universe was cool enough for atoms to form and from this moment on most of the radiation just streamed forth unimpeded.

This happened when the universe was a mere 300,000 years old.

So, can we see this relic of the ancient universe? The answer is yes! But before we look for it one thing must be kept in mind. The universe has been getting bigger and bigger and less and less dense. This implies that the average gravitational force is getting smaller with time. So the radiation, from the moment it no longer interacted with the newly formed atoms has been shifting from an environment where gravity's force is large to that where gravity is small and, using (again!) General Relativity, it must be red-shifted. In fact the *prediction* of General Relativity is that this radiation should be seen mostly as microwaves...and it *has* been seen. This prediction is not only of the existence of this relic radiation, but also how this radiation depends frequency . These predictions have been confirmed to great accuray (see Fig. 8.17). This ubiquitous sea of radiation that permeates the cosmos is called the *microwave background radiation*.

The microwave background radiation was created in approximately the same environment everywhere (remember that it came from an epoch in which everything was a very homogeneous hot mixture of nuclei and electrons) and because of this we expect it to look the same in every direction. This is precisely what happens, but, as it turns out, it is too much of a good thing: the microwave background radiation is the same everywhere to a precision of 0.1%, and understanding this presents problems, see Sect. 8.5.3.

But one can go even farther. Even though the microwave background radiation is very homogeneous, there *are* small deviations. These represent inhomogeneities in the universe at the time radiation and atoms stopped interacting strongly. These inhomogeneities provide a picture of the universe in its most tender infancy, see Fig. 8.18. As the universe expands and cools atoms will conglomerate into stars and stars into galaxies; the initial seeds for this process to start are these inhomogeneities. They correspond to regions where the matter was slightly mode dense than the average, and will, in the eons that follow, attract other matter to form the structures we see today.

It is very hard to explain the microwave background radiation by any theory other than the Big Bang. It represent one of its biggest successes.

8.4.3 Nucleosynthesis

The most abundant element in the universe is Hydrogen, the second most abundant element is Helium. A great success of the Big Bang theory is to be able to predict the relative amounts of these elements: after the universe

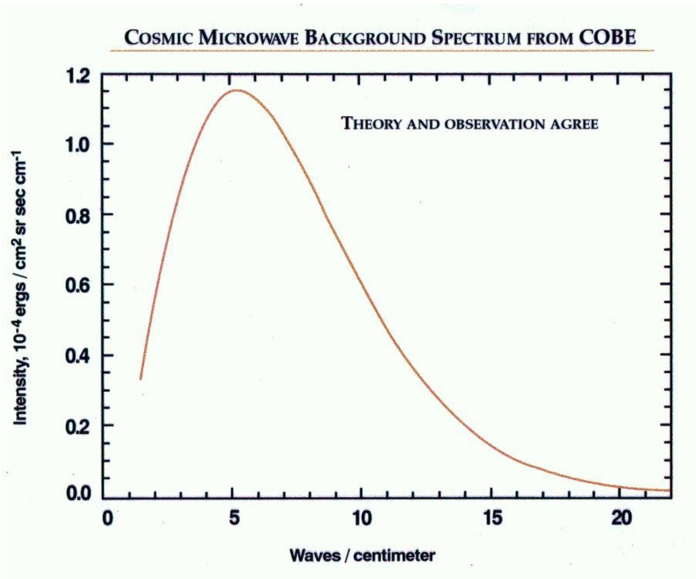

Figure 8.17: Radiation relics from the epoch shortly after the Big Bang. The horizontal axis corresponds to the frequency of the radiation, the vertical axis to the intensity. The measurements fall precisely on the curve.

cooled down sufficiently protons and neutrons were able, after a collision, to remain in the form of heavier atomic nuclei, in this manner Helium and Lithium were created, and also Deuterium (whose nucleus has one proton and one neutron). The universe was 1s old, its temperature was 10^{10o}K.

It was initially thought that *all* elements would be generated by the Big Bang, but this is not the case: even at the extreme temperatures available when Helium and Lithium nuclei were crated, this was not enough to smash two Helium nuclei to create something heavier, the creation of the remaining elements of the periodic table had to await the appearance of the first stars (see Sect. **??**). Deuterium and Lithium, while used up in stars through the nuclear reactions that make them shine (see Sect. **??**), are very rarely created by them . Whatever Deuterium and Lithium we see in nature was created about 15 billion years ago. Most of the Helium we observe (even though it is manufactured in stars) also came from that epoch.

The Big Bang theory predicts is the relative amounts of Helium and Lithium and Deuterium and Hydrogen. And the observations match the predictions; for example there are about 4 atoms of Hydrogen for each one of Helium. These same calcualations predict that there are 3 light neutrinos,

The Big Bang theory predicts is the relative amounts of Helium and Lithium and Deuterium and Hydrogen

The Big Bang theory predicts that there are 3 light neutrinos

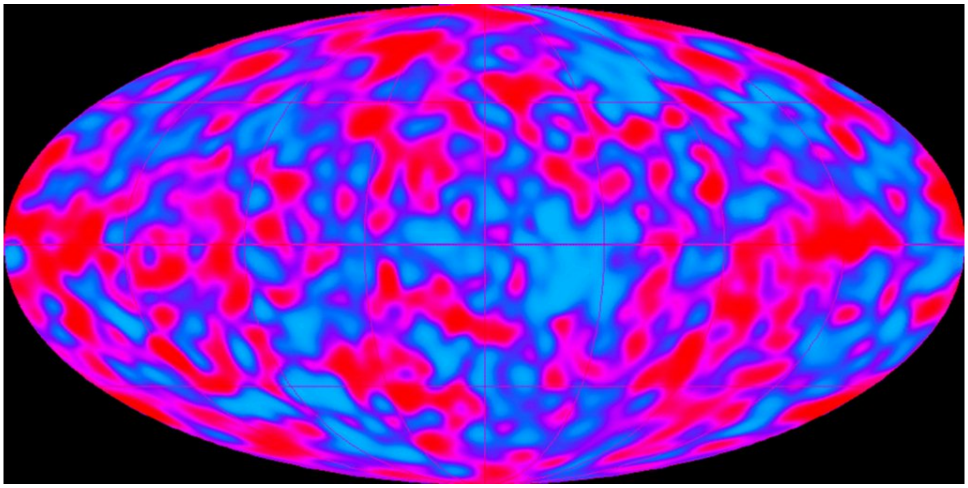

Figure 8.18: Inhomogeneities in the microwave background radiation. These give an idea of the way the universe looked shortly after the Big Bang.

again confirmed by observation.

Coupled with our understanding of stellar processes and evolution (Sects. **??** and **??**) we now understand the manner in which *all* elements in the periodic table were created. This is one of the most important predictions of modern cosmology.

We understand the manner in which *all* elements in the periodic table were created

8.5 At the cutting edge

Up to now all the results presented are well accepted and verified. There is little doubt that the General Theory of Relativity provides an excellent description of the universe at large, nor that the universe is currently expanding. Yet there are several puzzling results...

8.5.1 Dark matter

When considering the universe we observe only what we can see. Nonetheless there are strong indications that there is something more. Suppose you look at how stars in the outskirts of a galaxy move. Since gravity decreases with distance one would expect that the stars would slow down as the distance to the galaxy center increases, but this is *not* what is seen: the speed of these outlying stars appears to be constant (see Fig. 8.19). This is explained by assuming that the galaxy is in fact surrounded by a mass of matter which emits no or very little light, the so-called *dark matter*. In fact, calculations show that if this hypothesis is correct, this kind of matter is the main ingredient of galaxies, and perhaps the whole universe; an illustration of the "dark matter halo" surrounding a typical galaxy is given in Fig. 8.20

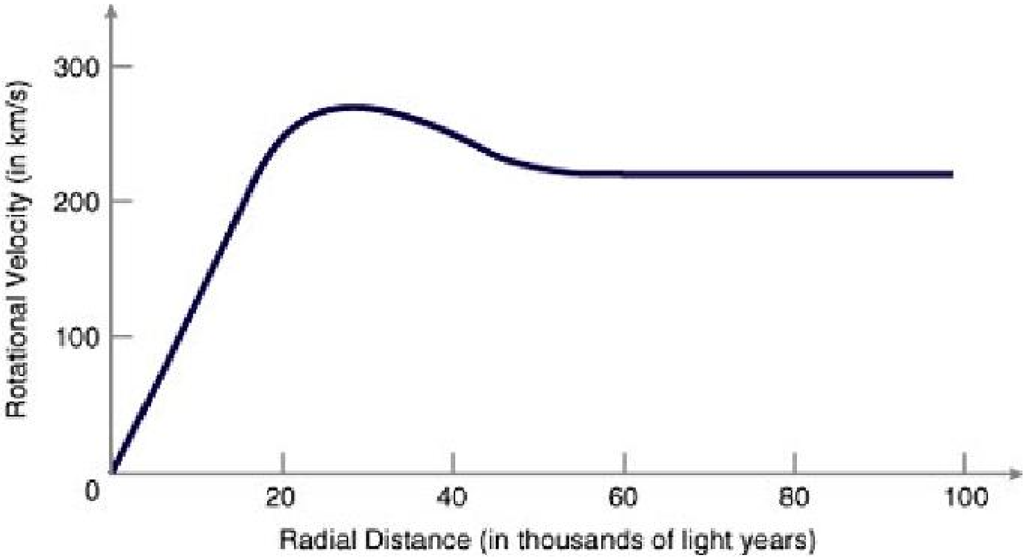

Figure 8.19: Rotation curve for stars in the Andromeda galaxy. The velocity becomes constant far away from the center suggesting the presence of dark matter.

What is this dark matter? No one knows! Is it perhaps a very large number of rocks, or planets? Is it something else? Or, maybe, is there a completely new effect which we interpret as dark matter while in reality there are new forces in action? The only recent answer is that there are strong indication that there are large numbers of planet-like objects in the vicinity of our galaxy. But these are not nearly enough to account for the whole effect. Many experiments are under way aiming at detecting the nature of dark matter (and it very existence).

Neutrinos

The early universe produced electromagnetic radiation which reaches us in the form of microwaves. This radiation was the result of the electromagnetic interactions among charged particles. There are, however, other types of interactions. We already met the gravitational interaction, and there are

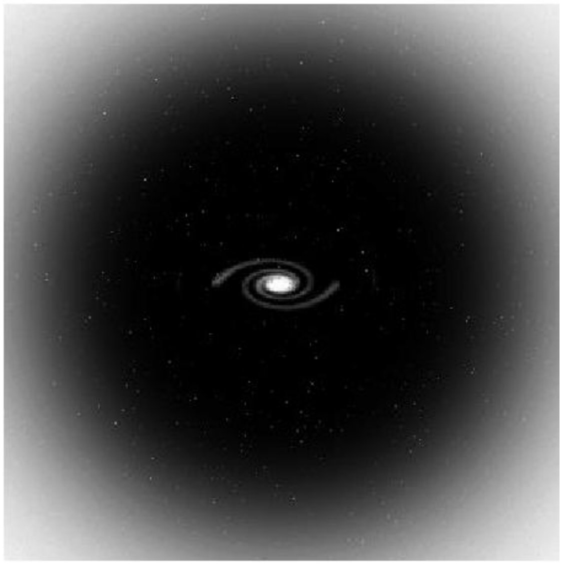

Figure 8.20: Illustration of the dark matter halo surrounding a typical galaxy.

two others called (again with a flair for words) the strong and the weak interactions.

Strong interactions are the ones responsible for nuclear forces between protons and neutrons (the constituents of atomic nuclei), and we will come back to them when we look at the evolution of a star (Sect. **??**). The remaining type, the weak forces, are experienced by *all* types of matter, but they are usually overwhelmed by the electromagnetic and strong forces because the weak interactions are, well, *weak!*

One is used to hear about electrons and protons and, perhaps to a lesser extent, neutrons. All these are constituents of atoms and atomic nuclei. But nature has a much richer population, and among its citizens one of the most intriguing are the neutrinos.

Neutrinos are very light particles [7] and experience *only* the weak interactions and it is because of this that they are rarely affected by other types of matter. Only in the densest of environments are neutrinos strongly disturbed. These occur in the center of neutron stars (Sec. **??**) or in the early universe. In this last case neutrinos were originally extremely energetic

---

[7] It had been assumed for a long time that they were massless, recent results however, indicate that neutrinos have a very small mass, of a billionth of a proton mass or less.

but, just as in the case of radiation, there came a time when the universe expanded to the point that the environment wasn't dense enough for the neutrinos to be affected by it. From that point on the neutrinos have been just cruising along, interacting only very rarely.

Initially these neutrinos lived in a very hot environment, which implies that each of them had a lot of energy *and* they were in a situation where very large gravitational forces were present. Nowadays they are in an environment where the gravitational forces are very weak. To understand what this implies consider the following analogy.

Imagine that you throw a ball up from the earth: initially the ball has a lot of kinetic energy, that is, energy due to its motion, but as it rises it slows down losing kinetic energy. Of course, this energy does not disappear, it is stored in potential energy (see Fig. 8.21). As the ball falls it will pick up speed so that when you catch it will be moving at the initial velocity (or close to it). In the same way the neutrinos in the present universe will have lost most of their kinetic energy.

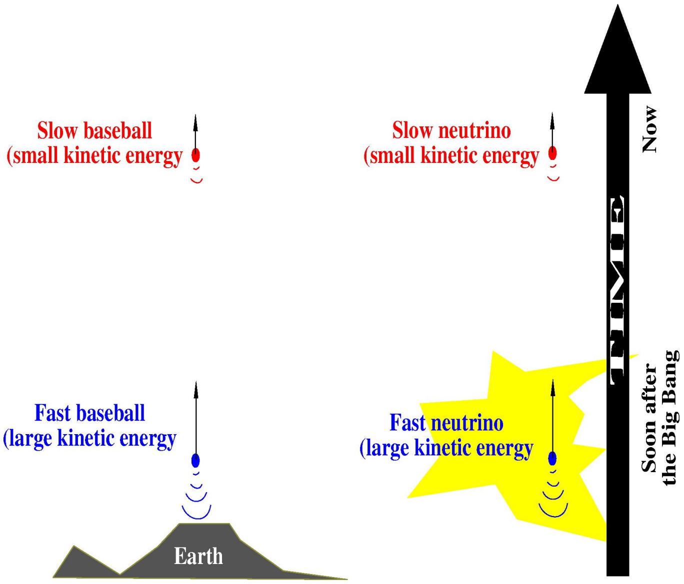

Figure 8.21: Neutrinos from the early universe have smaller kinetic energy now than in earlier epochs just as a baseball has lower kinetic energy the farther it is from Earth.

So another prediction of the Big Bang theory is that the universe is filled with neutrinos of very small kinetic energy. Unfortunately, out current technology is not sufficiently sophisticated to be able to detect them directly, but this might improve in the future.

8.5.2 The cosmological constant

When Einstein first studied the universe at large using the General Theory of Relativity he discovered that his equations predicted a universe which was either expanding or contracting, and this was contradicted with the best astronomical observations at the time. He then modified his equations to satisfy the observations. This modification corresponds to the assumption that the whole universe is permeated with a constant pressure (which in his case balanced the expansion yielding a steady universe). this universal pressure is called the *cosmological constant*

Though subsequently the data showed that the universe is in fact expanding and Einstein rejected the modification, on a philosophical basis the question still remains whether the *measured* cosmological constant is indeed zero (remember that on philosophical grounds Aristotle rejected heliocentrism: one must eventually back assumptions with observations). For many years the best value for the cosmological was assumed to be zero since no measurement gave positive indication to the contrary. Yet even a very small pressure can be important if it permeates the whole universe.

For many years the best value for the cosmological was assumed to be zero since no measurement gave positive indication to the contrary. Yet even a very small pressure can be important if it permeates the whole universe.

Recent measurements of the expansion rate of the universe (see Sect. 8.4.1) using type Ia supernovae (Sect. 8.4.1) favor an open universe with a small but non-zero cosmological constant. If these results are confirmed, Einstein's "blunder" will prove to be one more piece in the jigsaw of nature.:

8.5.3 Homogeneity and isotropy

One of the central simplifying assumptions of Einstein's cosmology is that, on average, the universe is the same in every direction (isotropy) and in every location (homogeneity). this does not mean, however, that the universe is a boring tapioca-like thing. The distribution of galaxies is far from smooth with most of them concentrated in relatively narrow sheets separated by large voids, see Fig. 8.22. The situation is reminiscent of a series of soap bubbles where the soapy water corresponds to the galaxies, the air inside the bubbles to the voids.

There are a few hipotheses which explain the origin of this type of structure. These must account not only for the voids, but also for the inhomogeneities in the comsic background radiation; and they must also predict a reasonable time-line for the development of galaxies. All these constraints are difficult to satisfy, making this an area of very active current research.

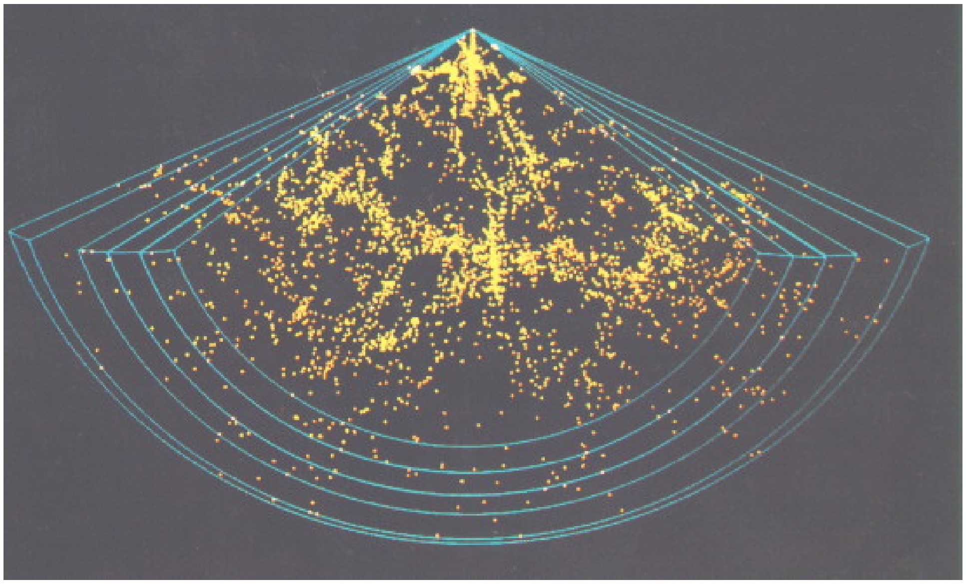

Figure 8.22: Large scale bubble-like structures in the universe. The image contains about 4000 galaxies each representing one luminous point.

Inflation

When we look at the microwave background radiation it looks the same in every direction, even from opposite sides of the sky, to a precision of 0.1%. Since they are so nicely correlated one would naturally assume that at some time all points in the observable universe were in close contact with each other, for otherwise it would be an unbelievable coincidence for all of them to look so much the same (at least through a microwave detector).

Now, a perfectly reasonable question is whether the Big Bang model has this property: will the Big Bang model predict not only the existence of the microwave background radiation, but also its exquisite uniformity? The answer is “yes” but only with additional assumptions.

This seems confusing: is the Big Bang theory to be modified and tuned every time a new piece of data comes along which does not agree with its predictions? Isn't this cheating? Doesn't this sound like Ptolemy adding epicycles every time things weren't quite accurate?

Fortunately this is not the case. The Big Bang theory determines the evolution of the universe *provided* the matter and energy content is known, *and* their behavior at very extreme conditions is well understood. The fact is, however, that we are not certain of all the matter and energy in the universe, nor do we know, for example, how they behave at temperatures above 10^{15} ^{o}K. Hence these "modifications" of the Big Bang theory correspond to different hypothesis of the behavior of matter at very high temperatures and densities, not of the general description provided by the General Theory of Relativity.

The simplest version of the Big Bang model which predicts a very uniforms microwave backgound goes by the way of *Inflation*. The idea is the following: the simplest way of getting uniform background radiation is if all the observable universe was in very close contact at an early time. Granted that, inflation provides a mechanism for increasing the size of this initially tiny region to the very large universe we see. Though mathematically involved what is assumed is that at a *very* time (about 10^{-35}s after the Big Bang) a new force comes into play which forces an exponential increase in the size of the universe (hence the name 'inflation'). After a fraction of a second this force is balanced by other interactions and the universe resumes a more dignified, if ponderous, expansion (see Fig. 8.23).

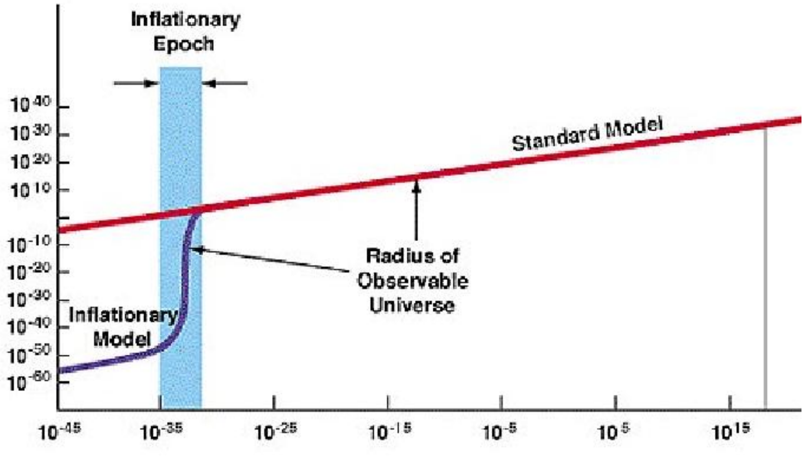

Figure 8.23: Time evolution of the size of the inflationary universe

One tantalizing conclusion derived from the inflationary hypothesis is that there are regions in the universe which we have not yet seen and which might look *very* different. Since no light has reached us from those regions we are currently unaware of their existence, only our inheritors will see the light coming from these distant reaches of the universe.

It is a challenge for current researchers to produce models that generate the intergalactic voids, yet with the *same* amount of dark matter required to understand the rotation of stars (Sect. 8.5.1) and using the inflation hypothesis such models actually exist. The corresponding computer simulations produce results such as the one shown in Fig. 8.24 which should be compared to the observations (Fig. 8.22).

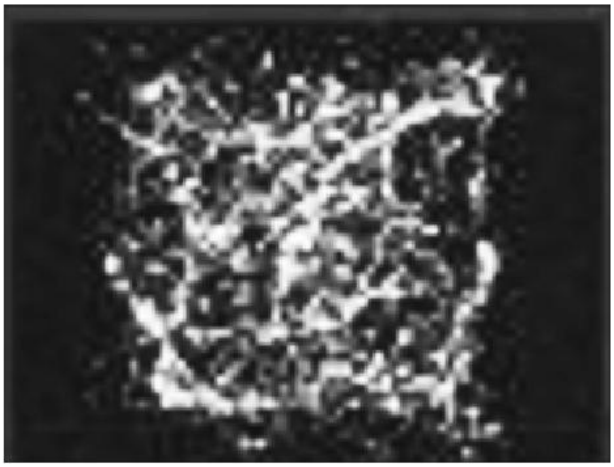

Figure 8.24: Simulation of the generation of structures in the universe assuming the presence of dark matter and an early epoch of inflation.

8.5.4 Summary

Though the General Theory of Relativity has produced a generous amount of verified predictions, its application to the universe at large has also generated a set of puzzles which, coupled to recent observations, are the topics of intense research. Whether there is a cosmological constant, whether the universe is filled with dark matter and the nature of this stuff and whether our current models of the universe are accurate enough to understand physics to the very earliest of times are issues currently addressed by researchers. The near future will provide more puzzles and some answers leading us, we hope, to a better understanding of the universe, our home.